AFIT/DSG/ENG/95-S-04

DTIC
SELECTED
AUG 1 4 1995
G

Optimal Inputs for System Identification

DISSERTATION
James M. Brown II
Captain

AFIT/DSG/ENG/95-S-04

19950811 057

Approved for public release; distribution unlimited

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the Department of Defense or the U. S. Government.

Accesion For		
NTIS CRA&I		☒
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

Optimal Inputs for System Identification

DISSERTATION

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

James M. Brown II, BSEE, MSEE

Captain

September, 1995

Optimal Inputs for System Identification

James M. Brown II, BSEE, MSEE

Captain

Approved:

Meir Pachter   11 July 1995

Robert N. Riggins   11 July 95

Peter S. Maybeck   11 July 95

Dennis W. Quinn   11 JULY 95

Robert A. Calico, Jr.

Dean

Acknowledgements

My beautiful wife, Mary, and my wonderful children, Andy, Alex, and Jamie, made this document possible. I could not have completed this research without their love and support. Since any list constitutes an exclusion list, I resist any attempt at thanking all those who had a part in the successful completion of my AFIT program. Having said that, I must single out the contributions made by my advisors, Meir Pachter and Bob Riggins. Their unflagging support and guidance got me going when I was down and accelerated me when I was up.

James M. Brown II

Table of Contents

Notation

Most of the notation used in this dissertation is taken directly from Maybeck [51].

Vectors, Matrices

Scalars. are denoted by upper or lower case letters in italic type.

Vectors. are denoted by lower case letters in boldface type, as the vector $\mathbf{x}$ made up of components x_i.

Matrices. are denoted by upper case letters in boldface type, as the matrix $\mathbf{A}$ made up of elements A_{ij} (ith row, jth column).

Random Vectors (Stochastic Processes), Realizations (Samples), and Dummy Variables

Random vectors. are set in boldface sans serif type, as x made up of scalar components x_i.

Realizations. of the random vector are set in boldface Roman type, as $\mathbf{x}$: $\mathsf{x}(\omega_i) = \mathbf{x}$.

Dummy variables. (for arguments of density or distribution functions, integrations, etc.) are denoted by the equivalent Greek letter, such as ξ being associated with x: e.g., $f_{\mathsf{x}}(\xi)$. The correspondences are (x, ξ), (y, ρ), (z, ζ).

Subscripts

> a : augmented c : continuous-time i, j, k, ℓ, m, n : indices into a vector,
> d : discrete-time t : true, truth model   matrix or sequence

Superscripts

$(\cdot)'$:   transpose (matrix)

$(\cdot)^{-1}$:   inverse (matrix or transform)

$\widehat{(\cdot)}$:   estimate

$(\cdot)^*$:   complement (set), complex conjugate, or conjugate transform

Matrix and Vector Relationships

$\mathbf{A} > \mathbf{B}$: $\mathbf{A}$ - $\mathbf{B}$ is positive definite

$\mathbf{A} \geq \mathbf{B}$: $\mathbf{A}$ - $\mathbf{B}$ is positive semidefinite

$\mathbf{x} \geq \mathbf{a}$:   component-wise, $x_1 \geq a_1, x_2 \geq a_2, \ldots, x_n \geq a_n$

Transforms and Operators

$\mathcal{F}\{\cdot\}$:   Discrete-Time Fourier transform

$\mathcal{L}\{\cdot\}$:   Laplace transform

$\mathcal{Z}\{\cdot\}$:   Z transform

$\mathcal{E}_{(\cdot)}\{\cdot\}$:   Expectation operator; the expectation is taken with respect to the subscript. The subscript may be omitted when there is no threat of ambiguity.

Sets

Sets are denoted by a blackboard font. Some commonly used sets are:

$\mathbb{R}$:   All real numbers

$\mathbb{R}^-$: $\{r \in \mathbb{R} \ni r \leq 0\}$

$\mathbb{R}^+$: $\{r \in \mathbb{R} \ni r \geq 0\}$

$\mathbb{Z}$:   All integers

$\mathbb{Z}^-$: $\{z \in \mathbb{Z} \ni z \leq 0\}$

$\mathbb{Z}^+$: $\{z \in \mathbb{Z} \ni z \geq 0\}$

$\mathbb{N}$:   All natural numbers

$\mathbb{C}$:   All complex numbers

List of Figures

List of Tables

List of Abbreviations

AFIT/DSG/ENG/95-S-04

Abstract

The derivation of the power spectral density of the optimal input for system identification is addressed in this research. Optimality is defined in information theoretic terms, with entropy quantifying the parameter-information content of the input and output measurement sequences pertaining to a discrete-time plant. The maximization of entropy is performed in the context of three different scenarios. First, the case in which the average output power of the plant is constrained is considered. Second, input average power is constrained. Finally, the optimization is carried out unconstrained, but with penalties applied to both the input and output powers. Although the focus of this research is the enhancement of the parameter identification potential of general System Identification algorithms, a new and efficient System Identification algorithm that employs Iterated Weighted Least Squares is derived. Experimental evidence is presented which clearly illustrates the superiority of this algorithm. Furthermore, experiments are documented which corroborate and validate the maximum-entropy-based theory for optimal input design presented in this dissertation.

Optimal Inputs for System Identification

I. Introduction

1.1 Motivation

With System Identification (ID) as a goal, we can expect perfect results if three sets of assumptions hold. First, the system must conform to the controls engineer's standard assumptions of linearity and time-invariance (LTI).[1] Second, we require noise-free measurements of the input and output histories. Third, we must provide excitation which is sufficient to make all the unknown parameters observable. The first set of assumptions are not overly restrictive. Although any true system is probably nonlinear and time-variant, the plant under inspection probably behaves linearly within some neighborhood about an operating point. Furthermore, most systems of interest exhibit slowly varying dynamics, effectively time-invariant for short periods. However, the assumption of perfect measurements is specious. We should *always* expect some level of corruption in the measurements through a combination of imperfect sensors or numerical quantization noise.

We can combat the effects of measurement noise by many techniques, all of which involve increasing some quantity related to the signal-to-noise ratio (SNR). One obvious solution involves increasing the magnitude of the signals to measure (assuming the noise is independent of the measured quantity). However, we usually do not wish to drive the system to extreme levels, motivating some sort of constraints on the input and output pairs. Furthermore, simply increasing the SNR is not a panacea for ID, as is shown in Section 6.5. Since we are faced with constraints on

---

[1]Of course, LTI assumptions are required only in the case in which we wish to find an LTI plant model. Since this research is focused on linear plants with constant parameters (with the possible exception of sudden changes, as in a failure), we restrict our attention to LTI models.

the magnitudes of the input and output, a logical approach involves increasing the signal levels in some judicious manner, with hopes of supplying the ID algorithm with the greatest possible amount of 'information' about the unknown system while limiting some metric describing the 'size' of the input and output.

Another motivation for quantifying and increasing the 'information' contained in the input/output pairs is exemplified by one problem associated with a closed-loop adaptive control system. Namely, a system under adaptive control can tend to lose observability of the parameters describing the plant under control. When this happens, the ID portion of the controller may generate very poor estimates of the math-model for the plant which leads to an incorrect controller. Improperly generated controls send the plant into wild excursions which, in turn, increase the parameter observability. As the observability increases, parameter estimates become reliable, facilitating a correct control sequence to recapture the desired plant trajectory. Again, once the plant is settled down by the controller, parameter observability drops, and the cycle is repeated. This well-known phenomenon, known as 'bursting', is shown in a transparent manner by the following example.

1.2 Bursting Example

1.2.1 System Description. The system we will use in this example is undamped, second order, discrete-time, linear, and time-invariant. We describe the system with the difference equation

$$y_{k+1} = ay_k - y_{k-1} + b(u_k + u_{k-1}) \tag{1}$$

with measurements given by

$$z_k = y_k + v_k \tag{2}$$

Alternatively, we describe the system with the transfer function:

$$H(z) = \frac{b(z+1)}{z^2 - az + 1} \tag{3}$$

2
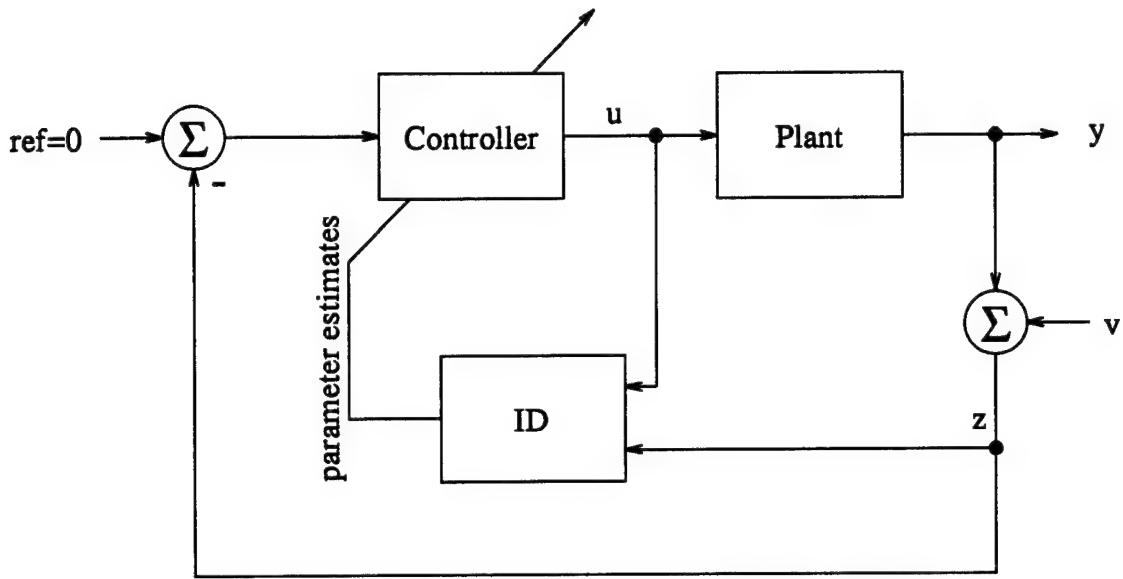
Figure 1. System Under Adaptive Control

where z is the forward shift operator. We wish to control the system with an adaptive controller set up as a regulator as shown in Figure 1.

1.2.2 Controller Design. Let us assume we wish for a type-one system for zero steady-state error. Thus, our desired open-loop transfer function is given by

$$GH = \frac{1}{z-1} \tag{4}$$

where G denotes the transfer function of the plant while H represents the transfer function of the controller. Applying the Guillemin-Truxal method[14], we solve for the controller, G

$$G(z) = \frac{1}{b} \frac{z^2 - az + 1}{z^2 - 1} \tag{5}$$

Equation (5) is based on *perfect* knowledge of the parameters, a and b. Since we are dealing with estimates of the parameters, we choose a controller design which does not try to cancel the system poles. Rather, the controller should perform a

3

near cancellation with zeros close to the poles but inside the unit circle. Thus, our desired controller is given by

$$G(z) = \frac{1}{\hat{b}} \frac{z^2 - 0.9\hat{a}z + 0.81}{z^2 - 1} \tag{6}$$

where $\hat{\ }$ indicates an estimate. Now, we must derive the algorithm to supply the estimates.

1.2.3 Estimation. From Equation (1) we have

$$\begin{bmatrix} y_k + y_{k-2} \\ y_{k-1} + y_{k-3} \end{bmatrix} = \begin{bmatrix} y_{k-1} & u_{k-1} + u_{k-2} \\ y_{k-2} & u_{k-2} + u_{k-3} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \tag{7}$$

which gives us

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_{k-1} & u_{k-1} + u_{k-2} \\ y_{k-2} & u_{k-2} + u_{k-3} \end{bmatrix}^{-1} \begin{bmatrix} y_k + y_{k-2} \\ y_{k-1} + y_{k-3} \end{bmatrix} \tag{8}$$

Equation (8) is based on perfect measurements. Since we are dealing with noise-corrupted quantities, we have

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} z_{k-1} & u_{k-1} + u_{k-2} \\ z_{k-2} & u_{k-2} + u_{k-3} \end{bmatrix}^{-1} \begin{bmatrix} z_k + z_{k-2} \\ z_{k-1} + z_{k-3} \end{bmatrix} \tag{9}$$

Careful inspection of Equations (8) and (9) begins to show us the source of the bursting problem. As the controller does its job, the output goes to zero, yielding a singular regressor matrix. Since our estimation scheme relies on the inversion of this matrix, the estimates breakdown.

1.2.4 Simulation Results. We see the effects of this break down of parameter observability in Figure 2. These plots were produced by a simulation of the system described previously, with $a = 1.8$, $b = 0.1$ and very low measurement noise ($\sigma = 10^{-8}$). Figure 2(a) clearly shows the bursting of the parameter estimates. Figure 2(b) displays the wildly oscillating output caused by improper control, brought on by incorrect parameter estimates. Also included in Figure 2(b) is the condition

4
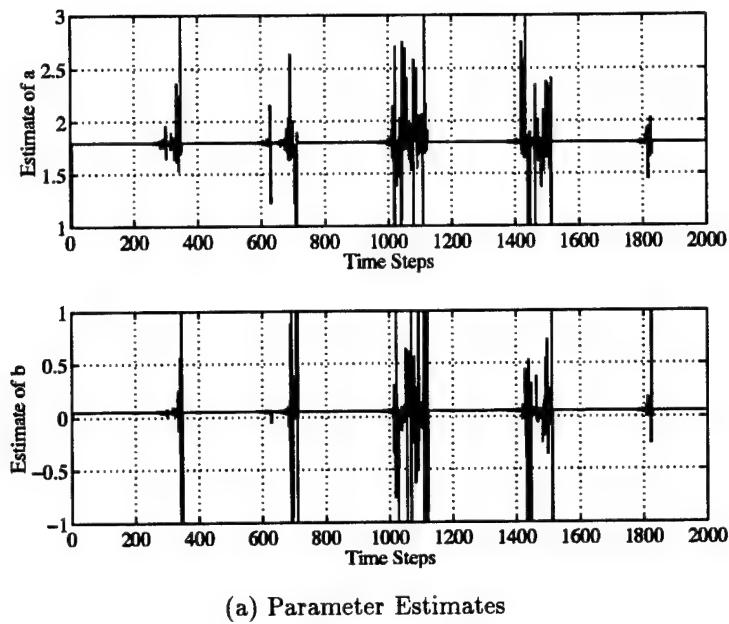
(a) Parameter Estimates
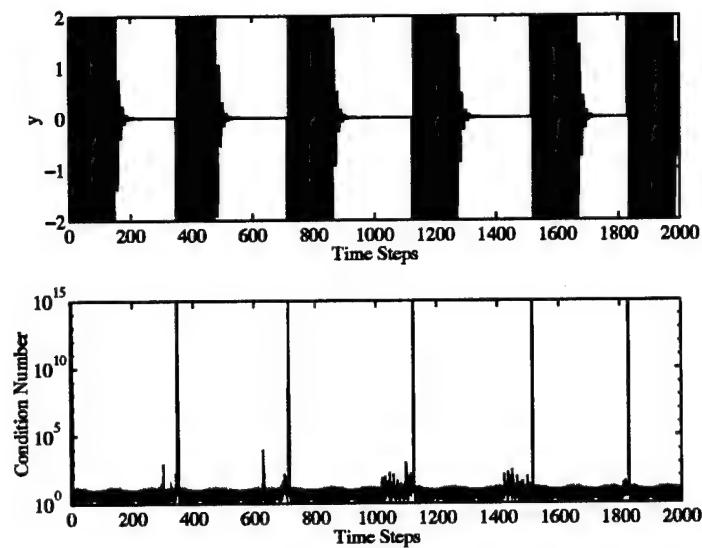
(b) Output and Condition Number

Figure 2.   Bursting Phenomenon Example: *These plots show the results of bursting under adaptive control. (a) shows the parameter estimates versus time. (b) shows the plant output and the regressor condition number versus time.*

5
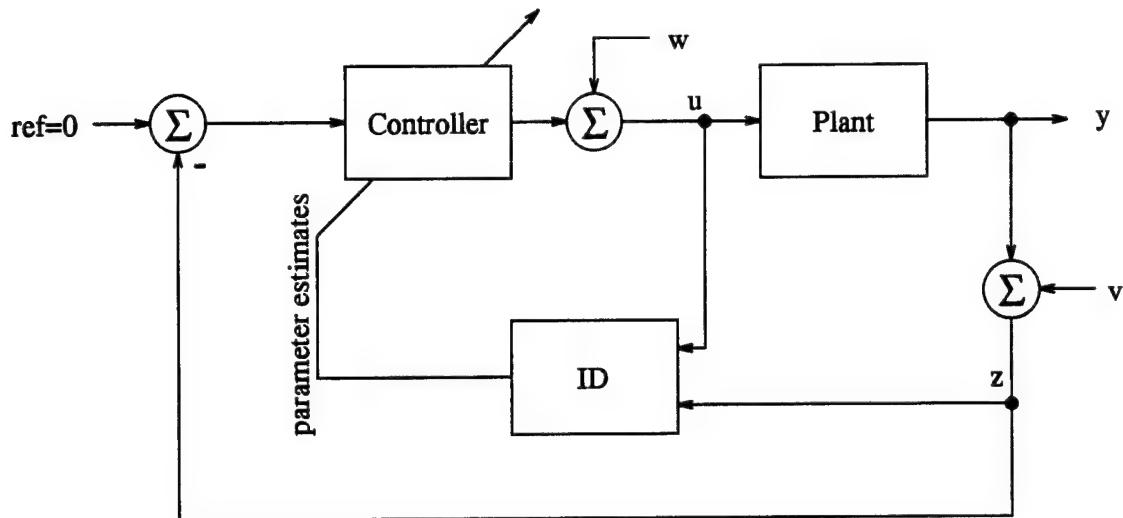
Figure 3. System Under Adaptive Control With an Auxiliary Input

number of the regressor matrix. The condition number is interesting to us in that it provides a measure of the singularity of the matrix to be inverted in the estimation scheme employed for this example. Comparing the plots, we see that as the system is regulated the condition number begins to rise. As the regressor matrix becomes ill-conditioned, the parameter estimates breakdown, beginning the cycle of unregulated output followed by better estimates, etc.

How can the bursting problem be alleviated? One solution is closely linked to the goal of this research. Namely, we inject an auxiliary input into the plant's input stream as shown in Figure 3. We see that this block diagram is identical to that shown in Figure 1, except we have added the auxiliary input **w**. If **w** is chosen properly, then the unknown parameters will always be observable (a condition known as *persistent excitation*, defined in [1, 2, 3, 42] and almost any other text devoted to System Identification).

A sufficient condition for persistent excitation is that the spectrum of the input be non-zero for at least as many points in frequency as the number of unknown parameters [44]. In order to ensure a persistently exciting input, we use a small-

6

variance white noise sequence for $\mathbf{w}$ ($\sigma_w^2 = 10^{-8}$). Thus, the input is persistently exciting of arbitrary order.

The results of the simulation with $\mathbf{w}$ injected are shown in Figure 4. We see in Figure 4(a) that the parameter estimates are now quite stable, with only sporadic excursions from the correct values. More importantly, the output (Figure 4)b)) is held tightly on zero, with no bursting. Thus, the adaptive regulator is enhanced by the auxiliary input.

1.2.5 Summary for Bursting Example. Bursting is a real concern in adaptive control. The example presented here clearly shows the effects of parameter observability break-down as an adaptive controller relies on parameter estimates which are derived from measurement-noise-corrupted outputs and inputs under conditions of poor excitation. Furthermore, we see that an auxiliary input injected into the plant's input can be designed to enhance excitation and facilitate parameter estimation, allowing an adaptive controller to regulate successfully the output of an unknown plant.

1.3 Proposed Adaptive Controller With Auxiliary Inputs

Given the problem of bursting in adaptively controlled systems, we envision a system which incorporates an input signal generator. The generator's function is to inject auxiliary inputs in order to keep parameter identifiability high.

Figure 5 shows a block diagram of the proposed adaptive control concept. That is, the proposed system is presented here as a conceptual tool, to illustrate the role System ID can play in adaptive control. The controller consists of the combination of prefilter and feed-forward compensator blocks which are designed to adapt to the changing plant. These control elements are controlled by the current best estimate of the plant parameters, thus achieving adaptivity. The estimates of the plant-model parameters are supplied by the ID block, along with an 'Excitation Level' signal.
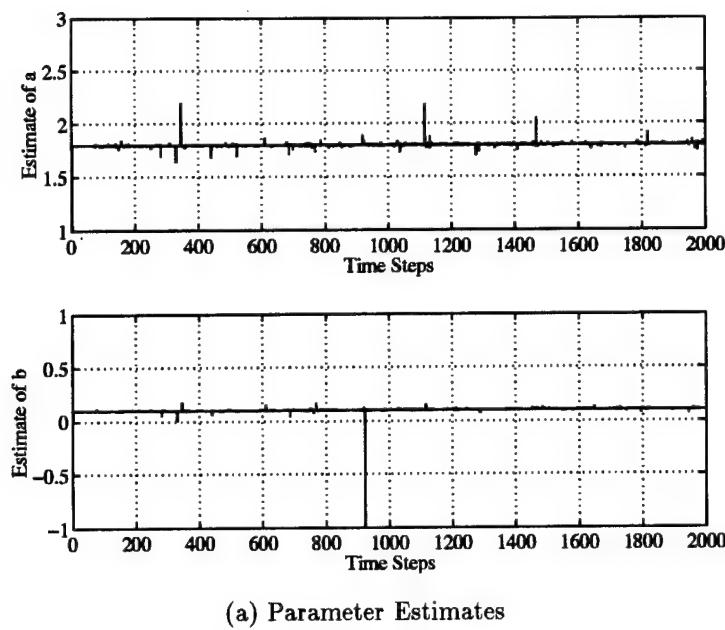
(a) Parameter Estimates
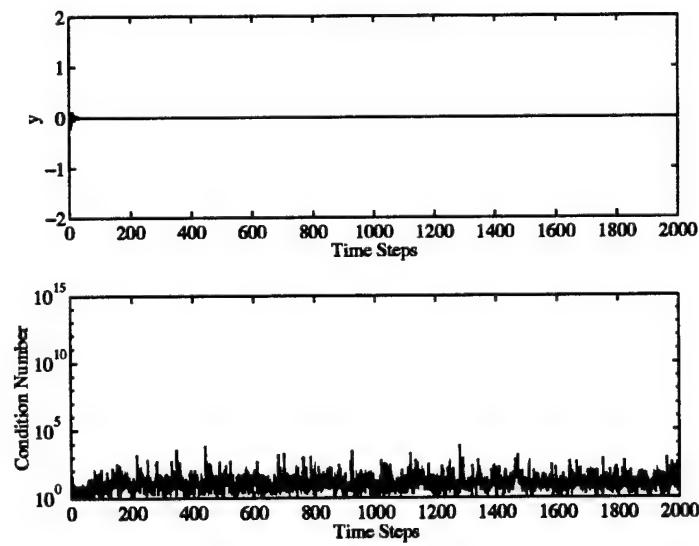
(b) Output and Condition Number

Figure 4. Alleviation of Bursting With Auxiliary Inputs: *These plots illustrate how bursting can be remedied by injecting an auxiliary input. The simulation which produced these plots is identical to that which produced Figure 2, but with an auxiliary input injected at the output of the controller. (a) shows the parameter estimates versus time. (b) shows the plant output and the regressor condition number versus time.*
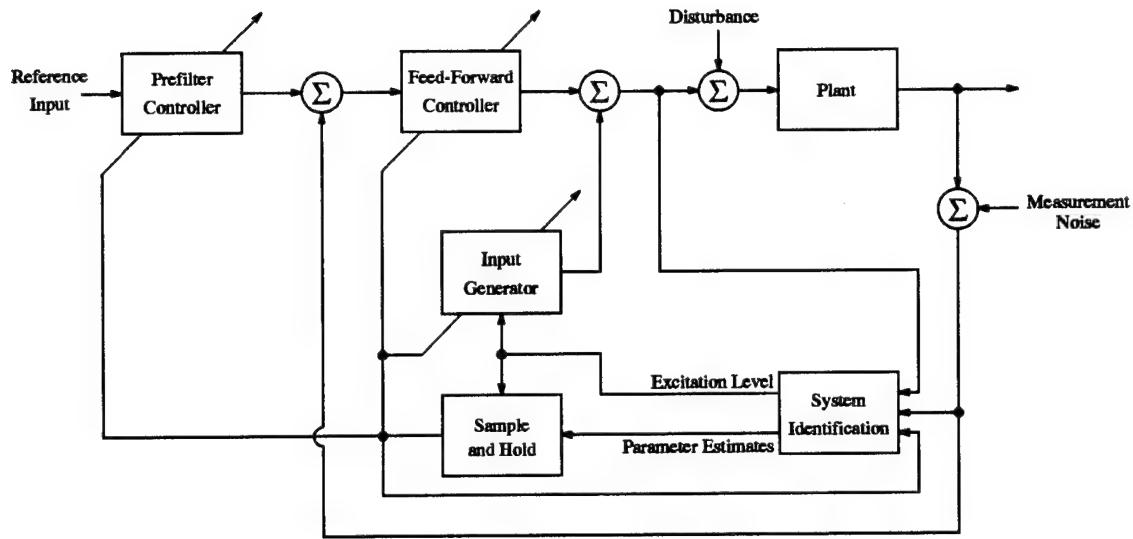
Figure 5. Block Diagram of a System Under Adaptive Control

The Excitation Level is a quantity which determines the confidence we have in the parameter estimates. For instance, the ID algorithm may generate an estimate of the parameter estimate error covariance, which could be used to generate the Excitation Level. This Excitation Level signal serves two purposes: First, this signal triggers a Sample and Hold which holds the last 'good' estimate of the parameters; Second, the Excitation Level triggers an Input Generator which produces probing inputs designed to enhance the identifiability of the parameters when such enhancement is required.

1.4 Summary of Current Literature

This research will focus on improving system identification with optimally (or sub-optimally) chosen inputs. Much research has been aimed at enhancing identification via designed inputs (also called *experiment design*). For example, Olmstead [67] derived a method for designing a constant feedback matrix which increases the information on parameters in the input/output pairs. Olmstead's work yields a matrix which modifies the output linearly, producing feedback inputs. In contrast, this

9

research will strive to design an *active* and *adaptive* input generator. The input generator uses knowledge of the history of the input, output, and parameter estimates to generate an input sequence which improves the ability of the system identification algorithm to extract parameter values. Thus, this research strives to improve the estimates produced by the ID algorithm.

Like Olmstead, many others have contributed to the field of input synthesis for identification. A representative set of works in this area is given by [1, 2, 3, 25, 28, 29, 45, 55, 56, 60, 61, 62, 67, 85, 86, 88, 92, 95, 96]. In many cases, the criterion for input optimality is linked directly to the ID technique employed. That is, an input is optimal only for a given identification algorithm.

In contrast, the work documented in this dissertation is aimed at maximizing the parameter information contained in the unknown plant's input/output pairs, independent of the chosen ID algorithm. Thus, the definition for optimality is not limited to any one form of identification, although a good ID algorithm is required to properly evaluate the ensuing ID performance. Many works address the concept of information, including [10, 13, 57, 68, 79, 80].

Although the goal of this research is to derive inputs for good ID, many sources related to the field of System Identification algorithms were reviewed for this research. In particular, [4, 18, 19, 21, 24, 26, 31, 32, 36, 39, 41, 42, 43, 44, 54, 58, 63, 64, 65, 69, 76, 84, 87, 90, 94] represent a small, but representative, portion of the work done in the field of System Identification. In the related field of fault, or change, detection references [6, 7, 8, 11, 12, 16, 22, 23, 33, 47, 48, 49, 50, 59, 66, 70, 71, 75, 89, 91, 97] present various techniques designed to recognize and identify a change in the parameters defining the mathematical model of the plant. The excitation is also important to change detection in that the input must excite the plant such that the parameters under scrutiny are observable. Obviously, good ID allows the estimation of parameters which are subject to change, and hence, also solves the fault detection and isolation problem.

Returning to experiment design, Mehra [55] presents a survey of the literature relevant to optimal inputs for system identification as of 1974. Mehra's paper tells us that the optimal input to minimize the trace of the estimation error covariance matrix is a white noise sequence (when the input is constrained by energy or amplitude). Furthermore, the optimal input in terms of minimizing the trace of the inverse of the Fisher Information Matrix (FIM) is bang-bang, again for amplitude-constrained inputs. An important contribution by Mehra reduces the steady-state problem to finite dimensions by proving that an input consisting of a finite number of frequencies can be found such that it yields the same information matrix as any other stationary input with equal power. Mehra's paper also points out an important consideration: Much research maximizes the trace (or weighted trace) of the FIM since this approach yields an easily solved quadratic problem. However, this criterion can produce a singular FIM. Thus, maximization of the trace of the FIM should be undertaken with care. Furthermore, Mehra's paper concentrates on constrained inputs whereas we shall consider three scenarios: constraints on the inputs; constraints on the outputs; and penalized inputs and outputs in an unconstrained framework.

Next, Gawthrop [24] discusses the problem of identification of systems in which a portion of the system is known. His paper allows us to identify systems which are polynomial, rather than linear, in the unknown parameters. For example, this technique allows one to identify an unknown system driven by known actuators. Unfortunately, assuming *known* actuators disallows us the chance to monitor the health of the actuators. This research will allow the identification of completely unknown plants, thereby directly allowing fault detection and isolation (FDI).

Aoki [2] discusses generating optimal input signals to enhance identifiability. However, Aoki's metric for optimization is the trace of the information matrix. As we mentioned previously, this optimization criterion can lead to a singular FIM.

Another paper by Mehra [56] presents several methods for solving the optimal input problem when using the trace of the FIM as the criterion and energy

11

constraints on the input. Mehra shows that the problem is quadratic, solvable via Riccati equations, the Ritz-Galerkin method, or the resolvent method. An example contained in the paper illustrates that the FIM tends to become singular; the condition number becomes large for the optimal input. Again, Mehra is dealing with constrained inputs, while we do not impose this restriction.

Ljung [42] treats the identification problem in the frequency domain, resulting in an infinite-dimensional problem. Although Ljung does not explicitly solve for an optimal input, the paper makes extensive use of the sequence driving the system. His criterion for good parameter estimates is that the discrete-time Fourier transform of the impulse response of the estimated transfer function be close to that of the actual system's transfer function. Ljung's approach is not suited to on-line ID, which is required for adaptive control.

Similar to Olmstead, Ng [61] seeks optimal inputs via an input/output feedback element. His approach is to minimize the determinant of the FIM, yielding so-called *D-optimal* estimates of the parameters. Ng concludes that input-constrained optimization requires no feedback (i.e. open-loop identification), while constraints on the output require inputs as a combination of feedback and open-loop inputs. Furthermore, he points out the paradox inherent in optimal input design – one must have knowledge of the parameters in order to design the inputs which optimize the parameter estimates. This research will address this very paradox through iteration and adaptation of the inputs, based on current parameter estimates.

Another paper by Ng [62] also uses the D-optimality criterion with constraints on the output power. Here, he concludes that the input is open-loop and constructed of a series of sinusoids. The sinusoidal inputs are found by solving a system of nonlinear equations. Again, the input design requires the use of the parameter values which we seek to estimate. As in Ljung's paper, this technique is suited only to off-line ID.

Later, Ng [60] worked around the problem of requiring knowledge of the parameters by using an iterative approach. He used the estimates from each iteration to recalculate the optimal inputs for the next iteration. The technique outlined in his paper is again based on D-optimality, but here, Ng presents the optimal input as a solution of a system of linear equations plus one polynomial equation. Ng's work is not suited to on-line ID since each iteration requires an entire input/output history.

Norton's paper [65] presents a departure from the standard stochastical approach in which noises and parameters are modeled as stochastic processes resulting in a search for the mean and covariance of the probability density functions. Rather, Nortons work revolves about a model incorporating *bounded* noise which is used to find *bounded regions* in parameter space in which the parameters lie. His approach allows the system identification to be used as a fault detection algorithm, signaling when the parameter estimates leave a predetermined region. Since the parameters are estimated to lie in a region, these estimates can be used in conjunction with control system synthesis techniques relying on structured uncertainty. Another appealing aspect of Norton's approach lies in the initialization of the algorithm; he allows the initial parameter region to consist of the entire parameter space. Thus, no prior knowledge of the parameters is required.

Incomplete knowledge of the parameters, coupled with an input of insufficient order (defined in the sequel), can cause a phenomenon known as *bursting*, described by Anderson in [1]. Anderson describes bursting as unstable, possibly oscillatory behavior between quiescent periods in the output of an identification algorithm based on output errors, or residuals. The inputs causing bursting are said to be lacking persistency of excitation. The bursting phenomenon is also known to manifest itself in adaptively controlled systems for which the input lacks persistency (or, by the terminology employed in this dissertation, lacks sufficient order). Our proposed input generator will maintain the proper excitation, thus eliminating the bursting problem.

13

Gevers [25] presents methods to design inputs for identification which require knowledge of the true parameters making up the system model. The important difference in this work is that Gevers approaches the inputs in terms of the intended use of the parameter estimates. He concludes that a feedback scheme is appropriate for most applications. As we see in much of the literature, the design of the optimal inputs requires the knowledge of the parameters.

The paradox involving the knowledge of the correct parameters inherent in the design of the optimal inputs continues throughout several other papers reviewed for this research. For example, see [4, 29]. Since we wish to apply system identification as a means to adaptive control and possibly fault detection, we desire to perform identification of the parameters with limited (or ideally, no) information about the parameters in the system model. Thus, we hope to probe the system based only on the history of parameter estimates (generated from a history of inputs and outputs) via an adaptive design of the input generator.

1.5 Contributions

The success of the research will contribute significantly to several different disciplines: system identification, adaptive control, and fault detection and isolation for reconfigurable control. The specific contributions are listed below:

1. Optimal input power spectral density is derived for any arbitrarily colored measurement noise and plant combination. Thus, we are not confined to a white measurement noise model. Furthermore, the derivations allow us to consider constraints on either the input power or the output power. Additionally, the optimal input is derived for the unconstrained case, with weighting applied to the input and output power.

2. An efficient and elegant ID algorithm is presented which correctly accounts for measurement noise and therefore yields good parameter estimates. The ID

14

algorithm adapts to changing parameter estimates in order to estimate properly the equation error covariance (i.e., pseudo-measurement noise, defined in Chapter III), which is determined by the measurement noise and the dynamics of the system under test. Experiments show the paramount importance of measurement noise handling in identification.

3. Theory and experiments are presented which underscore the nonlinearity of the System ID process. Although linear mathematical methods are used in the ID algorithm, said nonlinearity produces results which may surprise the engineer if he makes linear assumptions.

1.6 Organization of the Dissertation

The remainder of this dissertation is organized as follows. Chapter II covers deterministic estimation (modeling) in which all measured quantities are known perfectly, or the measurement noise is very small. The deterministic setting is used to gain insights into the identification problem and to establish some limits on parameter identifiability. With the observability of the parameters established, Chapter III complicates the problem with the addition of measurement noise. This formulation is realistic since, in the physical world, no measurement device is ideal; at the very least, we must deal with quantization error in the digital computer. Chapter III presents a method to weight properly a least squares estimate of the parameters. This Weighted Least Squares approach is finally used in the ID algorithm mentioned in the Contributions section above. Next, Chapter IV gives a detailed discussion of information theory and the link between information and ID. The information theoretic concepts presented here are used in the derivation of several types of optimal inputs: optimality in the face of input power constraints; optimality for output power constraints; and unconstrained optimal inputs with penalties applied to the input and output powers. In each case, the measurement noise is treated as arbitrary, allowing for colored noise. Experimental evidence is presented in Chapter V which

supports the theory presented in Chapter IV. More experiments are used in Chapter VI to investigate the practical limits of identification and to illustrate the effects of parameter errors on root locations and on frequency response. Furthermore, this chapter presents some surprising conclusions concerning the non-monotonicity of ID error with respect to signal-to-noise ratio. Finally, Chapter VII offers conclusions and recommendations for future work.

II. Deterministic Modeling

While the system described in Chapter I incorporates a *stochastic* model of the plant, we begin with a *deterministic* model in which all inputs and outputs are known 'perfectly'; the only noise affecting the problem is numeric. We have two reasons for initially ignoring physical noise. First, the mathematics are simpler, allowing us to develop some insights into the identification/modeling process. Without the complications involved with noisy signals, solvability conditions (akin to observability results in state estimation problems) can be established. Second, and more importantly, we can use the deterministic setting to establish limits on the identifiability of the parameters. In addition, the limitations on computation imposed by numerics – viz., the digital computer's internal noise – are gauged. The identification scheme must be able to extract the parameter values with no process and measurement noise if we hope to identify in the presence of noise.

2.1 Deterministic System Description

We consider a finite-dimensional, linear, time-invariant (LTI), single-input single-output (SISO), discrete-time system described as:

$$y_k = -\sum_{j=1}^{n} a_j y_{k-j} + \sum_{j=1}^{n} b_j u_{k-j} \tag{10}$$

Equation (10) may be rewritten as

$$y_k = \mathbf{h}'_{k-1}\boldsymbol{\theta} \tag{11}$$

where

$$
\begin{aligned}
\mathbf{h}'_{k-1} &= \begin{bmatrix} -y_{k-1} & -y_{k-2} & \cdots & -y_{k-n} & u_{k-1} & u_{k-2} & \cdots & u_{k-n} \end{bmatrix} \\
\boldsymbol{\theta}' &= \begin{bmatrix} a_1 & a_2 & \cdots & a_n & b_1 & b_2 & \cdots & b_n \end{bmatrix}
\end{aligned}
$$

17

Modeling entails the development of an algorithm that yields the $2n$ plant parameters (a_j and b_j, $j = 1, 2, \ldots, n$) with perfect knowledge of y_ℓ and u_ℓ $\forall \ell \in \{k - n, \ldots, k\}$.

2.2 Computation of the Parameters

Equation (10) outlines one possible method to describe a discrete-time system. If we assume the true system conforms to such a description, we reasonably expect to estimate successfully the parameters contained in $\boldsymbol{\theta}$. Two factors determine the amount of success we can expect in the estimation.

First, the form of the excitation (including initial conditions) is critical to the identification process. A degenerate example is zero input with zero initial conditions. Obviously, such an 'excitation' cannot provide information about the true system, for all linear systems will yield the same output of zero. This insight clearly illustrates that input affects the observability of the modeling problem, i.e., the identifiability of the parameters.

Second, we must consider the size of the proposed $\widehat{\boldsymbol{\theta}}$ vector versus the dimension of the actual $\boldsymbol{\theta}$ vector, viz., over-modeling and under-modeling. For example, if the proposed model's parameter vector is smaller than the true parameter vector, the model *cannot* completely describe the true system, since some modes of the true system can not be represented. On the other hand, a proposed model that is larger than the true system *might* describe the true system (with selected model parameters set to zero). However, we will see that over-specifying the order of the plant (viz., over-modeling) creates parameter-observability problems. Hence, we will show that over-modeling is a bad option.

Although a true system is probably infinite order, with a finite set of 'dominant' modes, the discussion in this chapter concerns concepts relevant to System ID; thus, we consider finite-order plants so we can explore the interaction between model and plant order.

18

2.3 Assumptions

Since this research specifically addresses excitation for identification, we consider only the case in which model order is known, as is oftentimes the case in electro/mechanical-dynamical systems encountered in some of engineering, e.g., in flight control. By assumption, we know the number of parameters $(2n)$, dispensing with the previous section's latter concern. Furthermore, we assume the system is bounded-input/bounded-output (BIBO) stable, allowing identification in the steady state. The first assumption is not overly restrictive; in science and engineering, we commonly have prior information about the number of dominant modes in a given system. For example, an aircraft's longitudinal motion typically is described by a pair of Phugoid poles and a pair of Short Period poles, plus known actuator poles, forming a model adequate for automatic control. The second assumption disallows a large set of real plants, but this research is concerned with steady-state excitation and experimentation, requiring a stable plant in the first place. With these assumptions, we can explore two different approaches to excitation for identification.

The two approaches explored in this chapter are *superposition of inputs* and *ensemble inputs*. The former, described in detail in Section 2.4, is the more realistic approach, wherein the input is described as a finite sum of complex exponentials. This scheme allows us to describe the output as another finite sum of complex exponentials and perform identification using linear regression on a history of outputs. The latter approach, ensemble inputs (Section 2.5), is a mathematical construct used to simplify the mathematics involved in proving some conjectures governing the requirements on the inputs for successful identification.

2.4 Approach I – Sum of Exponential Inputs

This approach involves the use of an input which is composed as a finite sum of distinct complex exponentials. We excite the system with such an input, wait for

steady-state behavior, and hope to calculate the parameter vector using a history of output measurements.

The general input used for this analysis is:

$$u_k = \sum_{i=1}^{N} B_i r_i^k, \ k \in \mathbb{Z}, \ B_i, r_i \in \mathbb{C} \tag{12}$$

Since the system is assumed to be linear, we can apply superposition to calculate the output. Furthermore, the system is in steady state, so the output is the sum of complex gains applied to the exponential inputs:

$$y_k = \sum_{i=1}^{N} A_i r_i^k, \ A_i \in \mathbb{C} \tag{13}$$

Now, considering a history of q outputs, we can construct the following linear regressions:

$$\begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-q+1} \end{bmatrix} = \begin{bmatrix} \mathbf{h}'_{k-1} \\ \mathbf{h}'_{k-2} \\ \vdots \\ \mathbf{h}'_{k-q} \end{bmatrix} \boldsymbol{\theta} \tag{14}$$

Or, in more compact notation, we express Equation (14) as

$$\mathbf{y}_k^{(q)} = \mathbf{H}_{k-1}^{(q)} \boldsymbol{\theta} \tag{15}$$

where

$$\mathbf{y}_k^{(q)} = \begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-q+1} \end{bmatrix}$$

and

$$\mathbf{H}_{k-1}^{(q)} = \begin{bmatrix} \mathbf{h}'_{k-1} \\ \mathbf{h}'_{k-2} \\ \vdots \\ \mathbf{h}'_{k-q} \end{bmatrix} = \begin{bmatrix} -y_{k-1} & \cdots & -y_{k-n} & u_{k-1} & \cdots & u_{k-n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -y_{k-q} & \cdots & -y_{k-n-q} & u_{k-q} & \cdots & u_{k-n-q} \end{bmatrix}$$

Now, if $\mathbf{H}_{k-1}^{(q)}$ is invertible, the parameter estimates $(\widehat{\boldsymbol{\theta}})$ are

$$\widehat{\boldsymbol{\theta}} = \left[\mathbf{H}_{k-1}^{(q)}\right]^{-1} \mathbf{y}_k^{(q)} \tag{16}$$

$\mathbf{H}_{k-1}^{(q)}$ is invertible if and only if it has full rank. Thus, identifiability of the parameters is contingent on $q \geq 2n$. Also, the $2n$ columns making up the matrix must be linearly independent. (We note that if $\mathbf{H}_{k-1}^{(q)}$ has full rank, $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$.) The conditions for linear independence are captured within the following discussion.

Definition 2.4.1 (Order of Excitation) *The order of excitation is equal to the minimum number of terms in the summation:*

$$u_k = \sum_i B_i r_i^k$$

required to describe the excitation (u_k), where $B_i, r_i \in \mathbb{C}$.

For example,

$$u_k = sin(k\theta\pi) = \left(\frac{1}{2j}\right)\left(e^{j\theta\pi}\right)^k - \left(\frac{1}{2j}\right)\left(e^{-j\theta\pi}\right)^k$$

is a second order excitation.

We now list some conjectures relating the order of the input to the identifiability of the system. The following conjectures are explored and confirmed with experimental evidence later in this chapter.

Conjecture 2.4.1 *The rank of $\mathbf{H}_{k-1}^{(q)}$ in steady state can be no greater than the order of the input. In other words, a necessary condition for identification in steady state is that the input order be at least as large as the number of unknown parameters.*

If we have access to the order of the plant, we strengthen the previous conjecture:

Conjecture 2.4.2 *If the order of the model and the order of the true system are equal, the window length (q) is greater than or equal to the number of parameters $(2n)$, and $\mathbf{H}_{k-1}^{(q)}$ is formed in steady state, then $\mathbf{H}_{k-1}^{(q)}$ has full rank if and only if the order of the input is greater than or equal to the number of unknown parameters.*

Note that these conjectures are proven indirectly via persistency of excitation arguments (e.g. see [44]).

2.5 Approach II – Ensemble Information

This approach is related to the former. Here, we conduct several experiments and compile the information from each into an 'ensemble information matrix' which we use to compute the parameter values.

As in the previous discussion, we consider exciting the system with complex exponentials. However, we do not sum the exponentials into one input and use a history of outputs. Rather, each input takes the form

$$u_{i_k} = r_i^k, \ r \in \mathbb{C}, \ i \in \mathbb{N}, \ k \in \mathbb{Z} \tag{17}$$

producing an output of the form

$$y_{i_k} = A_i r_i^k, \ A_i \in \mathbb{C} \tag{18}$$

Substituting Equations (17) and (18) into Equation (10), we have

$$A_i = \sum_{j=1}^{n} b_j r_i^{-j} - A_i \sum_{j=1}^{n} a_j r_i^{-j} \tag{19}$$

Thus, we have a linear equation involving the parameters we wish to compute. Since there are $2n$ parameters, we conduct the thought-experiment $2n$ times and construct the 'ensemble information matrix' equation:

$$
\begin{bmatrix}
A_1 r_1^n \\
\vdots \\
A_n r_n^n \\
A_{n+1} r_{n+1}^n \\
\vdots \\
A_{2n} r_{2n}^n
\end{bmatrix}
=
\begin{bmatrix}
1 & \cdots & r_1^{n-1} & -A_1 & \cdots & -A_1 r_1^{n-1} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & \cdots & r_n^{n-1} & -A_n & \cdots & -A_n r_n^{n-1} \\
1 & \cdots & r_{n+1}^{n-1} & -A_{n+1} & \cdots & -A_{n+1} r_{n+1}^{n-1} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & \cdots & r_{2n}^{n-1} & -A_{2n} & \cdots & -A_{2n} r_{2n}^{n-1}
\end{bmatrix}
\begin{bmatrix}
b_n \\
\vdots \\
b_1 \\
a_n \\
\vdots \\
a_1
\end{bmatrix}
\tag{20}
$$

or, for notational ease,

$$\mathbf{\Gamma} = \mathbf{\Psi}\theta \tag{21}$$

Note that the parameter vector is rearranged in order to achieve the particular form for the $\mathbf{\Psi}$ given in Equation (20).

Inspecting Equation (20), we see that the 'ensemble information matrix', $\mathbf{\Psi}$, can be partitioned into the following sub-matrices:

$$\mathbf{\Psi} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{\Lambda}_1\mathbf{V}_1 \\ \mathbf{V}_2 & \mathbf{\Lambda}_2\mathbf{V}_2 \end{bmatrix} \tag{22}$$

where $\mathbf{V}_1$ and $\mathbf{V}_2$ are *Vandermonde* matrices, having the form

$$\mathbf{V} = \begin{bmatrix} r_1 & r_2 & \cdots & r_m \\ r_1^2 & r_2^2 & \cdots & r_m^2 \\ \vdots & \vdots & \ddots & \vdots \\ r_1^m & r_2^m & \cdots & r_m^m \end{bmatrix}$$

and $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are diagonal matrices. This special structure of $\mathbf{\Psi}$ should aid in deriving the conditions for which the system is identifiable.

2.6 Equivalence of Ensemble and Summation Approaches

We are interested in identifying the parameters in the system defined by:

$$y_k = \sum_{j=1}^{n} a_j y_{k-j} + \sum_{j=1}^{n} b_j u_{k-j} \tag{23}$$

Consider an input consisting of a sum of $2n$ complex exponentials:

$$u_k = \sum_{i=1}^{2n} r_i^k \quad r_i \in \mathbb{C} \tag{24}$$

In steady state, the output is given by:

$$y_k = \sum_{i=1}^{2n} A_i r_i^k \quad A_i \in \mathbb{C} \tag{25}$$

23

We have $2n$ parameters, so consider $2n$ measurements of the output:

$$\begin{bmatrix} y_k \\ \vdots \\ y_{k-2n+1} \end{bmatrix} = \mathbf{H}_{k-1}^{(2n)} \boldsymbol{\theta} \tag{26}$$

where

$$\mathbf{H}_{k-1}^{(2n)} = \begin{bmatrix} \sum_{i=1}^{2n} r_i^{k-n} & \cdots & \sum_{i=1}^{2n} r_i^{k-1} & -\sum_{i=1}^{2n} A_i r_i^{k-n} & \cdots & -\sum_{i=1}^{2n} A_i r_i^{k-1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{2n} r_i^{k-3n+1} & \cdots & \sum_{i=1}^{2n} r_i^{k-2n} & -\sum_{i=1}^{2n} A_i r_i^{k-3n+1} & \cdots & -\sum_{i=1}^{2n} A_i r_i^{k-2n} \end{bmatrix}$$

$$\boldsymbol{\theta} \triangleq \begin{bmatrix} b_n \\ \vdots \\ b_1 \\ a_n \\ \vdots \\ a_1 \end{bmatrix}$$

In order to solve for $\boldsymbol{\theta}$, we must have nonsingular $\mathbf{H}_{k-1}^{(2n)}$. Now, $\mathbf{H}_{k-1}^{(2n)}$ is a *Casorati* [37] matrix for the set of sequences

$$\mathbb{S} = \left\{ \sum_{i=1}^{2n} r_i^{k-1}, \ \sum_{i=1}^{2n} r_i^{k-2}, \ \cdots, \ \sum_{i=1}^{2n} r_i^{k-n}, \ \sum_{i=1}^{2n} -A_i r_i^{k-1}, \ \cdots, \ \sum_{i=1}^{2n} -A_i r_i^{k-n} \right\}$$

So, $\mathbf{H}_{k-1}^{(2n)}$ is nonsingular if and only if $\mathbb{S}$ is a linearly independent set of sequences. Thus, write a *different Casorati* matrix for the same set:

$$\mathbf{K} \triangleq \begin{bmatrix} \sum_{i=1}^{2n} r_i & \sum_{i=1}^{2n} r_i^2 & \cdots & \sum_{i=1}^{2n} r_i^n & -\sum_{i=1}^{2n} A_i r_i & \cdots & -\sum_{i=1}^{2n} A_i r_i^n \\ \sum_{i=1}^{2n} r_i^2 & \sum_{i=1}^{2n} r_i^3 & \cdots & \sum_{i=1}^{2n} r_i^{n+1} & -\sum_{i=1}^{2n} A_i r_i^2 & \cdots & -\sum_{i=1}^{2n} A_i r_i^{n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{2n} r_i^{2n} & \sum_{i=1}^{2n} r_i^{2n+1} & \cdots & \sum_{i=1}^{2n} r_i^{3n-1} & -\sum_{i=1}^{2n} A_i r_i^{2n} & \cdots & -\sum_{i=1}^{2n} A_i r_i^{3n-1} \end{bmatrix} \tag{27}$$

24

So $\mathbf{K}$ is nonsingular $\Leftrightarrow$ $\mathbb{S}$ is linearly independent $\Leftrightarrow$ $\mathbf{H}_{k-1}^{(2n)}$ is invertible. This particular form for $\mathbf{K}$ is chosen with hindsight. We will see the justification in the following discussion.

Now, consider conducting $2n$ experiments with the exponentials used in the previous development. Recall from Equation (20)

$$
\mathbf{\Psi} = \begin{bmatrix}
1 & \cdots & r_1^{n-1} & -A_1 & \cdots & -A_1 r_1^{n-1} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & \cdots & r_n^{n-1} & -A_n & \cdots & -A_n r_n^{n-1} \\
1 & \cdots & r_{n+1}^{n-1} & -A_{n+1} & \cdots & -A_{n+1} r_{n+1}^{n-1} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & \cdots & r_{2n}^{n-1} & -A_{2n} & \cdots & -A_{2n} r_{2n}^{n-1}
\end{bmatrix} \tag{28}
$$

Given this 'ensemble' approach, the parameters are identifiable if $\mathbf{\Psi}$ is nonsingular.

Now, consider the matrix product $\mathbf{V\Psi}$ where $\mathbf{V}$ is the *Vandermonde* matrix given by

$$
\mathbf{V} = \begin{bmatrix}
r_1 & r_2 & \cdots & r_{2n} \\
r_1^2 & r_2^2 & \cdots & r_{2n}^2 \\
\vdots & \vdots & \ddots & \vdots \\
r_1^{2n} & r_2^{2n} & \cdots & r_{2n}^{2n}
\end{bmatrix} \tag{29}
$$

Combining Equations (28) and (29), we have

$$
\mathbf{V\Psi} = \begin{bmatrix}
\sum_{i=1}^{2n} r_i & \sum_{i=1}^{2n} r_i^2 & \cdots & \sum_{i=1}^{2n} r_i^n & -\sum_{i=1}^{2n} A_i r_i & \cdots & -\sum_{i=1}^{2n} A_i r_i^n \\
\sum_{i=1}^{2n} r_i^2 & \sum_{i=1}^{2n} r_i^3 & \cdots & \sum_{i=1}^{2n} r_i^{n+1} & -\sum_{i=1}^{2n} A_i r_i^2 & \cdots & -\sum_{i=1}^{2n} A_i r_i^{n+1} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^{2n} r_i^{2n} & \sum_{i=1}^{2n} r_i^{2n+1} & \cdots & \sum_{i=1}^{2n} r_i^{3n-1} & -\sum_{i=1}^{2n} A_i r_i^{2n} & \cdots & -\sum_{i=1}^{2n} A_i r_i^{3n-1}
\end{bmatrix} \tag{30}
$$

Thus, $\mathbf{V\Psi} = \mathbf{K}$. Now, $\mathbf{V}$ is a *Vandermonde* matrix, and thus, is guaranteed to be nonsingular (providing the r's making up the matrix are distinct). It follows that $\mathbf{\Psi}$

25

is nonsingular *if and only if* **H** is invertible; we have established an inexorable link between the two experimental approaches.

Now we have a simpler form with which to deal in the proof of Conjecture 2.4.2. If we prove the link between the rank of **Ψ** and input order, then the same link is established between **H** and input order. Clearly, **Ψ** is a simpler matrix to use in the proof.

2.7 Invertibility of **Ψ**

The previous section establishes the link between the summation approach and the ensemble approach. Now, we must prove the invertibility of the ensemble information matrix, **Ψ**.

Before considering the general case, we can establish the nonsingularity of **Ψ** for special cases. If we assume the order of the plant under investigation is known, we can calculate the determinant of **Ψ** for the proper number of exponential inputs. Equation (10) gives the time-domain representation of the true system under scrutiny. However, since we are dealing with steady-state excitation, it is more convenient to use the transfer function given by:

$$\mathbf{G}(\zeta) = \frac{g \prod_{j=1}^{n-1}(\zeta - z_j)}{\prod_{j=1}^{n}(\zeta - p_j)} \tag{31}$$

where the difference-equation parameters are absorbed in g, z_j, and p_j. In steady state, the system behaves as a complex gain. In other words

$$u_{i_k} = r_i^k \quad \Rightarrow \quad y_{i_k} = A_i r_i^k \tag{32}$$

where

$$A_i = \mathbf{G}(\zeta)|_{\zeta = r_i} = \frac{g \prod_{j=1}^{n-1}(r_i - z_j)}{\prod_{j=1}^{n}(r_i - p_j)} \tag{33}$$

26

where ζ is the forward shift operator and $\{z_j\}$ and $\{p_j\}$ are the poles and zeros of the transfer function, respectively.

2.7.1 First Order Plant Example.   For a general first order plant, we have two unknowns, g and p_1:

$$\mathbf{G}(\zeta) = \frac{g}{\zeta - p_1}$$

From Equations (28) and (33)

$$\mathbf{\Psi} = \begin{bmatrix} 1 & \frac{-g}{r_1 - p_1} \\ 1 & \frac{-g}{r_2 - p_1} \end{bmatrix}$$

We can easily compute the determinant of $\mathbf{\Psi}$:

$$|\mathbf{\Psi}| = g\frac{(r_2 - r_1)}{(r_1 - p_1)(r_2 - p_1)} \tag{34}$$

So, $\mathbf{\Psi}$ is invertible if and only if $g \neq 0$ and $r_1 \neq r_2$. In other words, sufficient conditions for identifiability of the first-order system are second order input and a non-trivial system, as predicted by Conjectures 2.4.1 and 2.4.2.   ∎

2.7.2 Second Order Plant Case.   A general second order plant is described by four unknown parameters (g, z_1, p_1, and p_2):

$$\mathbf{G}(\zeta) = \frac{g(\zeta - z_1)}{(\zeta - p_1)(\zeta - p_2)}$$

Again, Equations (28) and (33) give us

$$\mathbf{\Psi} = \begin{bmatrix} 1 & r_1 & \frac{-g(r_1 - z_1)}{(r_1 - p_1)(r_1 - p_2)} & \frac{-g(r_1 - z_1)}{(r_1 - p_1)(r_1 - p_2)}r_1 \\ 1 & r_2 & \frac{-g(r_2 - z_1)}{(r_2 - p_1)(r_2 - p_2)} & \frac{-g(r_2 - z_1)}{(r_2 - p_1)(r_2 - p_2)}r_2 \\ 1 & r_3 & \frac{-g(r_3 - z_1)}{(r_3 - p_1)(r_3 - p_2)} & \frac{-g(r_3 - z_1)}{(r_3 - p_1)(r_3 - p_2)}r_3 \\ 1 & r_4 & \frac{-g(r_4 - z_1)}{(r_4 - p_1)(r_4 - p_2)} & \frac{-g(r_4 - z_1)}{(r_4 - p_1)(r_4 - p_2)}r_4 \end{bmatrix}$$

The algebra involved in the computation of the determinant is cumbersome, so we present the explicit and elegant result, calculated by Mathematica [93]:

$$|\Psi| =$$
$$g^2 \frac{(r_4 - r_3)(r_4 - r_2)(r_4 - r_1)(r_3 - r_2)(r_3 - r_1)(r_2 - r_1)(p_1 - z_1)(p_2 - z_1)}{(r_1 - p_1)(r_1 - p_2)(r_2 - p_1)(r_2 - p_2)(r_3 - p_1)(r_3 - p_2)(r_4 - p_1)(r_4 - p_2)} \quad (35)$$

2.7.2.1 Discussion of the Second Order Case: We see results similar to the first order case. Here Ψ is nonsingular for $r_i \neq r_j$ $(i \neq j)$, $g \neq 0$, and no pole/zero cancellation. Thus, with these minimal restrictions on the plant, the identifiability of the parameters is determined by the order of the input; a second order plant requires a fourth order input for identifiability.

2.7.3 Third Order Plant Case. Finally, we investigate the Ψ regressor matrix for a general third order plant (with six unknown parameters).

$$\mathbf{G}(\zeta) = \frac{g(\zeta - z_1)(\zeta - z_2)}{(\zeta - p_1)(\zeta - p_2)(\zeta - p_3)}$$

Again, by Equations (28) and (33)

$$\Psi = \begin{bmatrix} 1 & r_1 & r_1^2 & -A_1 & -A_1 r_1 & -A_1 r_1^2 \\ 1 & r_2 & r_2^2 & -A_2 & -A_2 r_2 & -A_2 r_2^2 \\ 1 & r_3 & r_3^2 & -A_3 & -A_3 r_3 & -A_3 r_3^2 \\ 1 & r_4 & r_4^2 & -A_4 & -A_4 r_4 & -A_4 r_4^2 \\ 1 & r_5 & r_5^2 & -A_5 & -A_5 r_5 & -A_5 r_5^2 \\ 1 & r_6 & r_6^2 & -A_6 & -A_6 r_6 & -A_6 r_6^2 \end{bmatrix}$$

where

$$A_i = \frac{g(r_i - z_1)(r_i - z_2)}{(r_i - p_1)(r_i - p_2)(r_i - p_3)}$$

Once again, the algebra involved is difficult. However, Mathematica [93] computes the determinant:

$$|\Psi| = -g^3 \frac{\left\{ \begin{array}{c} (r_6 - r_5)(r_6 - r_4)(r_6 - r_3)(r_6 - r_2)(r_6 - r_1)\times \\ (r_5 - r_4)(r_5 - r_3)(r_5 - r_2)(r_5 - r_1)(r_4 - r_3)\times \\ (r_4 - r_2)(r_4 - r_1)(r_3 - r_2)(r_3 - r_1)(r_2 - r_1)\times \\ (p_1 - z_1)(p_2 - z_1)(p_3 - z_1)(p_1 - z_2)(p_2 - z_2)(p_3 - z_2) \end{array} \right\}}{\left\{ \begin{array}{c} (r_1 - p_1)(r_1 - p_2)(r_1 - p_3)(r_2 - p_1)(r_2 - p_2)(r_2 - p_3)\times \\ (r_3 - p_1)(r_3 - p_2)(r_3 - p_3)(r_4 - p_1)(r_4 - p_2)(r_4 - p_3)\times \\ (r_5 - p_1)(r_5 - p_2)(r_5 - p_3)(r_6 - p_1)(r_6 - p_2)(r_6 - p_3) \end{array} \right\}} \tag{36}$$

We see that the third order case continues the pattern. The determinant of Ψ is zero if and only if the set of r's is not distinct, or the true plant has pole/zero cancellations, or the plant is trivial ($g = 0$). Again, the identifiability of the parameters is determined by the input order.

2.7.4   nth Order Plant.   The three special cases indicate a possible pattern for the determinant of Ψ, which we state as the following Conjecture:

Conjecture 2.7.1 *If the Ψ regressor matrix is constructed as outlined in this section, the determinant, $|\Psi|$, is given by*

$$|\Psi| = g^n \frac{\left\{ \prod_{i=1}^{2n-1} \prod_{j=i+1}^{2n} (r_j - r_i) \right\} \left\{ \prod_{i=1}^{n} \prod_{j=1}^{n-1} (p_i - z_j) \right\}}{\prod_{i=1}^{2n} \prod_{j=1}^{n} (r_i - p_j)}$$

We leave this conjecture unproven. However, accepting its validity, we see in a very transparent way that an input of order $2n$ is necessary and sufficient for identification of the parameters describing an nth order plant with no pole/zero cancellations. Furthermore, the following discussion introduces an identification theorem which holds for any LTI plant.

2.8 Effect of Pole/Zero Cancellations

The three important special cases investigated in Section 2.7, along with the generalization in Conjecture 2.7.1, indicate that identifiability of the parameters is partially dependent on the location of the poles and zeros of the transfer function. This dependence is stated as the following theorem:

Theorem 2.8.1 *Regardless of excitation, the regressor matrix* $\mathbf{H}$ *is invertible only if the model upon which* $\mathbf{H}$ *is based lacks pole/zero cancellations. That is, a necessary condition for identifiability is a minimum-order model.*

Proof: Consider the $\mathbf{H}$ matrix based on an nth order model. Then the columns of $\mathbf{H}$ consist of the past n inputs and the past n outputs ($u_{k-1} \ldots u_{k-n}$ and $y_{k-1} \ldots y_{k-n}$). Now, pole/zero cancellation effectively reduces the system order to something less than n. Thus, the last output (y_{k-1}) can be expressed as a linear combination of $y_{k-2} \ldots y_{k-n}$ and $u_{k-1} \ldots u_{k-n}$, yielding a rank-deficient $\mathbf{H}$ matrix. ∎

Since one of our basic assumptions is the lack of pole/zero cancellations in the plant, how do we encounter these cancellations? The answer lies in the chosen order of the model we wish to identify. For example, consider a second order true system. Expanding Equation (31) we have

$$\mathbf{G}(\zeta) = \frac{b_1\zeta + b_2}{\zeta^2 - a_1\zeta - a_2} \tag{37}$$

But, if the proposed model were third order, Equation (31) would yield

$$\widehat{\mathbf{G}(\zeta)} = \frac{\hat{b}_1\zeta^2 + \hat{b}_2\zeta + \hat{b}_3}{\zeta^3 - \hat{a}_1\zeta^2 - \hat{a}_2\zeta - \hat{a}_3} \tag{38}$$

Now, for $\mathbf{G} = \widehat{\mathbf{G}} \; \forall \; \zeta \in \mathbb{C}$ we require $\hat{b}_3 = \hat{a}_3 = 0$ which reduces Equation (38) to

$$\widehat{\mathbf{G}(\zeta)} = \frac{\hat{b}_1\zeta^2 + \hat{b}_2\zeta}{\zeta^3 - \hat{a}_1\zeta^2 - \hat{a}_2\zeta} = \frac{\hat{b}_1\zeta + \hat{b}_2}{\zeta^2 - \hat{a}_1\zeta - \hat{a}_2} \tag{39}$$

where the pole and zero at the origin are canceled. This procedure is extended easily to plants of arbitrary order. Any proposed model of order greater than that of

the true system yields pole/zero cancellations at the origin. Hence, simple-minded over-modeling for determining the true order of the plant is not advised.

2.9 Conclusions

So far, we have explored some of the relationships linking input to identifiability of the parameters given perfect knowledge of the input and output. Several examples illustrate that the order of the input must be at least as great as the number of parameters to estimate. Furthermore, we maintain that this relationship between input order and identifiability extends to plants of arbitrary order.

Now that we have gained these insights in a deterministic setting, we turn in Chapter III to a *stochastic* setting, wherein the measurements on the plant are corrupted by noise. Obviously, the deterministically derived identifiability conditions are necessary conditions for identifiability in the case with measurement noise.

III. Stochastic Estimation

Chapter II treats parameter estimation, given perfect knowledge of inputs and outputs (i.e. deterministic estimation, or modeling). Now, we complicate the problem with the inevitable inclusion of measurement and process noise. That is, we now consider the problem of *System Identification.*

The pertinent stochastic system model is given by the underlying dynamics

$$y_k = -\sum_{j=1}^{n} a_j y_{k-j} + \sum_{j=1}^{n} b_j u_{k-j} + \sum_{j=1}^{n} g_j w_{k-j} \tag{40}$$

and the measurements

$$z_k = y_k + v_k \tag{41}$$

where y is an internal variable, u is a deterministic input, w is a realization of random process noise, and v is a realization of the random measurement noise. The respective process and measurement noise quantities, w_k and v_k, are independent white sequences with Gaussian distribution and statistics given by:

$$\left. \begin{aligned} \mathcal{E}\{w_k\} &= 0 \\ \mathcal{E}\{w_j w_k\} &= \sigma_w^2 \delta_{jk} \\ \mathcal{E}\{v_k\} &= 0 \\ \mathcal{E}\{v_j v_k\} &= \sigma_v^2 \delta_{jk} \\ \mathcal{E}\{v_j w_k\} &= 0 \end{aligned} \right\} \quad \forall\, j, k \in \mathbb{Z} \tag{42}$$

where δ_{jk} is the Kronecker delta function.

The importance for ID of the concise dynamic stochastic model given in Equations (40) – (42) cannot be overemphasized. In particular, notice the difference between process noise and measurement noise and the correct modeling in Equation 41) of the physical measurement process. A distinction is clearly drawn between the underlying ideal system state y and its available measurement z. We shall see that the scalar components making up the equation error vector associated with a

history of measurements are cross-correlated by the dynamics of the system. In contrast, the process-noise terms in the equation error vector are cross-correlated with correlation terms affected only by the moving average coefficients (g_j) defining the model for the process noise. Since the measurement-noise contribution to the cross-correlation terms involves the a_j terms (which are unknown), this causes significant complications for System Identification.

Recalling Equation (15) in Chapter II we have

$$\mathbf{y}_k = \mathbf{H}_{k-1}^{(q)} \boldsymbol{\theta} \tag{43}$$

with one very important difference. Since we are dealing with noise corrupted data, we must use the *measurements* of the output; the y data is *not* directly available experimentally – this problem does not exist in the deterministic formulation. Thus,

$$\mathbf{z}_k^{(q)} = \mathbf{H}_{k-1}^{(q)} \boldsymbol{\theta} + \mathbf{n}_k^{(q)} \tag{44}$$

where now

$$\mathbf{z}_k^{(q)} = \begin{bmatrix} z_k \\ z_{k-1} \\ \vdots \\ z_{k-(q-1)} \end{bmatrix} \tag{45}$$

$$\mathbf{H}_{k-1}^{(q)} = \begin{bmatrix} -z_{k-1} & \cdots & -z_{k-n} & u_{k-1} & \cdots & u_{k-n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -z_{k-q} & \cdots & -z_{k-n-q+1)} & u_{k-q} & \cdots & u_{k-n-q+1} \end{bmatrix} \tag{46}$$

and

$$\boldsymbol{\theta}' = \begin{bmatrix} a_1 & \vdots & a_n & b_1 & \vdots & b_n \end{bmatrix} \tag{47}$$

and the vector $\mathbf{n}_k^{(q)}$ denotes the *equation error*, explained in the following sections. We shall see that $\mathbf{n}$ is not equal to a simple vectorization of v_k, due to correlation

induced by the dynamics of the system. Note also that the entries of $\mathbf{H}$ and $\mathbf{z}$ consist of the recorded measurements, rather than the "clean" signal y.

Assuming $q \geq 2n$, the straight pseudo-inverse to calculate an estimate for the parameter vector is often employed [53], i.e., the Least Squares (LS) estimate is

$$\widehat{\boldsymbol{\theta}} = \left[\mathbf{H}_{k-1}^{(q)}{}' \mathbf{H}_{k-1}^{(q)} \right]^{-1} \mathbf{H}_{k-1}^{(q)}{}' \mathbf{z}_{k}^{(q)} \tag{48}$$

Thus, Equation (48) calculates the parameter estimates via an unweighted pseudo-inverse. $\widehat{\boldsymbol{\theta}}$ is referred to as the Least Squares estimate. However, the rigorous minimum-error-variance estimate is given by the minimum-variance estimate, which uses a properly *weighted* pseudo-inverse:

$$\widehat{\boldsymbol{\theta}}_{MV} = \left[\mathbf{H}_{k-1}^{(q)}{}' \mathbf{R}^{-1} \mathbf{H}_{k-1}^{(q)} \right]^{-1} \mathbf{H}_{k-1}^{(q)}{}' \mathbf{R}^{-1} \mathbf{z}_{k}^{(q)} \tag{49}$$

It has been shown [79] that the $\mathbf{R}$ matrix in Equation (49) which minimizes the estimation error covariance is given by the covariance of the 'equation error.' Thus, we must first compute this covariance matrix.

3.1 $\mathbf{R}$ *With No Process Noise*

If we assume that we have no process noise, then $\mathbf{R}$ is only affected by the measurement noise, $\mathbf{v}$. We may be tempted to think that $\mathbf{R}$ is simply a diagonal matrix with σ_v^2 along the diagonal. In fact, some estimation schemes make this error. For example, the ARX command in the Matlab® ID toolbox is based on this assumption. However, using the ARX command for the ID of the dynamical system described in Equations (40) – (42), a practice often encountered, is erroneous. We will see that $\mathbf{R}$ is more complex than a simple diagonal matrix.

We begin with an example. Consider a second-order system given by the dynamics

$$y_k = -a_1 y_{k-1} - a_2 y_{k-2} + b_1 u_{k-1} + b_2 u_{k-2} \tag{50}$$

and the measurements

$$z_k = y_k + v_k \tag{51}$$

Combining Equations (50) and (51), we have

$$
\begin{aligned}
z_k &= -a_1(z_{k-1} - v_{k-1}) - a_2(z_{k-2} - v_{k-2}) + b_1 u_{k-1} + b_2 u_{k-2} + v_k \\
&= -a_1 z_{k-1} - a_2 z_{k-2} + b_1 u_{k-1} + b_2 u_{k-2} + v_k + a_1 v_{k-1} + a_2 v_{k-2} \tag{52}
\end{aligned}
$$

If we use a window length of $q = 5$, we have

$$
\begin{bmatrix} z_k \\ z_{k-1} \\ z_{k-2} \\ z_{k-3} \\ z_{k-4} \end{bmatrix}
=
\begin{bmatrix}
-z_{k-1} & -z_{k-2} & u_{k-1} & u_{k-2} \\
-z_{k-2} & -z_{k-3} & u_{k-2} & u_{k-3} \\
-z_{k-3} & -z_{k-4} & u_{k-3} & u_{k-4} \\
-z_{k-4} & -z_{k-5} & u_{k-4} & u_{k-5} \\
-z_{k-5} & -z_{k-6} & u_{k-5} & u_{k-6}
\end{bmatrix}
\begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix}
$$
$$
+
\begin{bmatrix}
v_k + a_1 v_{k-1} + a_2 v_{k-2} \\
v_{k-1} + a_1 v_{k-2} + a_2 v_{k-3} \\
v_{k-2} + a_1 v_{k-3} + a_2 v_{k-4} \\
v_{k-3} + a_1 v_{k-4} + a_2 v_{k-5} \\
v_{k-4} + a_1 v_{k-5} + a_2 v_{k-6}
\end{bmatrix} \tag{53}
$$

or

$$\mathbf{z}_k^{(5)} = \mathbf{H}_{k-1}^{(5)} \boldsymbol{\theta} + \mathbf{n}_k^{(5)} \tag{54}$$

where $q = 5$ is chosen arbitrarily for this example; note that q must be greater than or equal to four for four unknowns. Furthermore, we note that the regressor

matrix, $\mathbf{H}$, is populated entirely with measured quantities. Given the assumptions on v (Equation (42)), and Equations (53) and (54), we have

$$\mathbf{R} = \mathcal{E}\left\{\mathbf{n}_k^{(q)}\mathbf{n}_k^{(q)'}\right\} =$$
$$\begin{bmatrix} 1+a_1^2+a_2^2 & a_1a_2+a_1 & a_2 & 0 & 0 \\ a_1a_2+a_1 & 1+a_1^2+a_2^2 & a_1a_2+a_1 & a_2 & 0 \\ a_2 & a_1a_2+a_1 & 1+a_1^2+a_2^2 & a_1a_2+a_1 & a_2 \\ 0 & a_2 & a_1a_2+a_1 & 1+a_1^2+a_2^2 & a_1a_2+a_1 \\ 0 & 0 & a_2 & a_1a_2+a_1 & 1+a_1^2+a_2^2 \end{bmatrix}\sigma_v^2 \qquad (55)$$

3.2 $\mathbf{R}$ With Process Noise

Now, we further complicate the problem with the addition of process noise. The new system model is given by

$$y_k = -a_1y_{k-1} - a_2y_{k-2} + b_1u_k + b_2u_{k-1} + g_1w_k + g_2w_{k-1} \qquad (56)$$

$$z_{k+1} = y_{k+1} + v_{k+1} \qquad (57)$$

Isolating the autoregressive parameters, the moving average parameters, and the noise quantities, we have

$$\begin{aligned} z_{k+1} = \ & -a_1(z_k - v_k) - a_2(z_{k-1} - v_{k-1}) \\ & + \ b_1u_k + b_2u_{k-1} \\ & + \ v_{k+1} + g_1w_k + g_2w_{k-1} \end{aligned} \qquad (58)$$

A little rearranging gives

$$\begin{aligned} z_{k+1} = \ & (-a_1z_k - a_2z_{k-1} + b_1u_k + b_2u_{k-1}) \\ & + \ (v_{k+1} - a_1v_k - a_2v_{k-1}) \\ & + \ g_1w_k + g_2w_{k-1} \end{aligned} \qquad (59)$$

36

Again, using a window length of $q = 5$:

$$
\begin{bmatrix} z_k \\ z_{k-1} \\ z_{k-2} \\ z_{k-3} \\ z_{k-4} \end{bmatrix} = \begin{bmatrix} z_{k-1} & z_{k-2} & u_{k-1} & u_{k-2} \\ z_{k-2} & z_{k-3} & u_{k-2} & u_{k-3} \\ z_{k-3} & z_{k-4} & u_{k-3} & u_{k-4} \\ z_{k-4} & z_{k-5} & u_{k-4} & u_{k-5} \\ z_{k-5} & z_{k-6} & u_{k-5} & u_{k-6} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix}
$$

$$
+ \begin{bmatrix} v_k + a_1 v_{k-1} + a_2 v_{k-2} \\ v_{k-1} + a_1 v_{k-2} + a_2 v_{k-3} \\ v_{k-2} + a_1 v_{k-3} + a_2 v_{k-4} \\ v_{k-3} + a_1 v_{k-4} + a_2 v_{k-5} \\ v_{k-4} + a_1 v_{k-5} + a_2 v_{k-6} \end{bmatrix} + \begin{bmatrix} g_1 w_{k-1} + g_2 w_{k-2} \\ g_1 w_{k-2} + g_2 w_{k-3} \\ g_1 w_{k-3} + g_2 w_{k-4} \\ g_1 w_{k-4} + g_2 w_{k-5} \\ g_1 w_{k-5} + g_2 w_{k-6} \end{bmatrix} \tag{60}
$$

or

$$
\begin{aligned}
\mathbf{z}_k^{(5)} &= \mathbf{H}_{k-1}^{(5)} \boldsymbol{\theta} + \mathbf{n}_k^{(5)} \\
&= \mathbf{H}_{k-1}^{(5)} \boldsymbol{\theta} + \mathbf{v}_k^{(5)} + \mathbf{w}_{k-1}^{(5)}
\end{aligned} \tag{61}
$$

The $\mathbf{R}$ matrix must include *both* $\mathbf{v}_k^{(5)}$ and $\mathbf{w}_{k-1}^{(5)}$ terms:

$$
\begin{aligned}
\mathbf{R} &= \mathcal{E}\left\{ \left(\mathbf{v}_k^{(5)} + \mathbf{w}_{k-1}^{(5)} \right) \left(\mathbf{v}_k^{(5)} + \mathbf{w}_{k-1}^{(5)} \right)' \right\} \\
&= \mathcal{E}\left\{ \mathbf{v}_k^{(5)} \mathbf{v}_k^{(5)'} \right\} + \mathcal{E}\left\{ \mathbf{v}_k^{(5)} \mathbf{w}_{k-1}^{(5)}{}' \right\} + \mathcal{E}\left\{ \mathbf{w}_{k-1}^{(5)} \mathbf{v}_k^{(5)'} \right\} + \mathcal{E}\left\{ \mathbf{w}_{k-1}^{(5)} \mathbf{w}_{k-1}^{(5)}{}' \right\} \\
&= \mathcal{E}\left\{ \mathbf{v}_k^{(5)} \mathbf{v}_k^{(5)'} \right\} + \mathcal{E}\left\{ \mathbf{w}_{k-1}^{(5)} \mathbf{w}_{k-1}^{(5)}{}' \right\}
\end{aligned} \tag{62}
$$

where

$$
\mathcal{E}\left\{ \mathbf{v}_k^{(5)} \mathbf{v}_k^{(5)'} \right\} =
$$
$$
\begin{bmatrix} 1 + a_1^2 + a_2^2 & a_1 a_2 + a_1 & a_2 & 0 & 0 \\ a_1 a_2 + a_1 & 1 + a_1^2 + a_2^2 & a_1 a_2 + a_1 & a_2 & 0 \\ a_2 & a_1 a_2 + a_1 & 1 + a_1^2 + a_2^2 & a_1 a_2 + a_1 & a_2 \\ 0 & a_2 & a_1 a_2 + a_1 & 1 + a_1^2 + a_2^2 & a_1 a_2 + a_1 \\ 0 & 0 & a_2 & a_1 a_2 + a_1 & 1 + a_1^2 + a_2^2 \end{bmatrix} \sigma_v^2 \tag{63}
$$

37

and

$$\mathcal{E}\left\{\mathbf{w}_{k-1}^{(5)}\mathbf{w}_{k-1}^{(5)}{}'\right\} =$$
$$\begin{bmatrix} g_1^2 + g_2^2 & g_1g_2 & 0 & 0 & 0 \\ g_1g_2 & g_1^2 + g_2^2 & g_1g_2 & 0 & 0 \\ 0 & g_1g_2 & g_1^2 + g_2^2 & g_1g_2 & 0 \\ 0 & 0 & g_1g_2 & g_1^2 + g_2^2 & g_1g_2 \\ 0 & 0 & 0 & g_1g_2 & g_1^2 + g_2^2 \end{bmatrix} \sigma_w^2 \tag{64}$$

The matrices defining $\mathbf{R}$ are generalized to

$$\mathcal{E}\left\{\mathbf{v}_k^{(q)}\mathbf{v}_k^{(q)'}\right\} = \begin{bmatrix} \nu_0 & \nu_1 & \cdots & \nu_{n-1} & \nu_n & 0 & \cdots & 0 \\ \nu_1 & \nu_0 & \nu_1 & & \nu_{n-1} & \ddots & \ddots & \vdots \\ \vdots & \nu_1 & \ddots & \ddots & & \ddots & \ddots & 0 \\ \nu_{n-1} & & \ddots & \ddots & \ddots & & \nu_{n-1} & \nu_n \\ \nu_n & \nu_{n-1} & & \ddots & \ddots & \ddots & & \nu_{n-1} \\ 0 & \ddots & \ddots & & & \ddots & \ddots & \nu_1 & \vdots \\ \vdots & \ddots & \ddots & \nu_{n-1} & & \nu_1 & \nu_0 & \nu_1 \\ 0 & \cdots & 0 & \nu_n & \nu_{n-1} & \cdots & \nu_1 & \nu_0 \end{bmatrix} \sigma_v^2 \tag{65}$$

$$\mathcal{E}\left\{\mathbf{w}_k^{(q)}\mathbf{w}_k^{(q)'}\right\} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} & \gamma_{n-1} & 0 & \cdots & 0 \\ \gamma_1 & \gamma_0 & \gamma_1 & & \gamma_{n-2} & \ddots & \ddots & \vdots \\ \vdots & \gamma_1 & \ddots & \ddots & & \ddots & \ddots & 0 \\ \gamma_{n-2} & & \ddots & \ddots & \ddots & & \gamma_{n-2} & \gamma_{n-1} \\ \gamma_{n-1} & \gamma_{n-2} & & \ddots & \ddots & \ddots & & \gamma_{n-2} \\ 0 & \ddots & \ddots & & & \ddots & \ddots & \gamma_1 & \vdots \\ \vdots & \ddots & \ddots & \gamma_{n-2} & & \gamma_1 & \gamma_0 & \gamma_1 \\ 0 & \cdots & 0 & \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix} \sigma_w^2 \tag{66}$$

where

$$\nu_0 = 1 + \sum_{j=1}^{n} a_j^2 \tag{67}$$

$$\nu_\ell = a_\ell + \sum_{j=\ell+1}^{n} a_j a_{j-\ell}, \quad \ell = 1, 2, \ldots, n \tag{68}$$

$$\gamma_0 = \sum_{j=1}^{n} g_j^2 \tag{69}$$

$$\gamma_\ell = \sum_{j=\ell+1}^{n} g_j g_{j-\ell}, \quad \ell = 1, 2, \ldots, n \tag{70}$$

$$\tag{71}$$

We note that both matrices have some nice properties. They are symmetric and Toeplitz matrices [27]. Therefore, each $q \times q$ matrix is fully defined by a small number of parameters, reducing storage requirements. Furthermore, efficient algorithms exist for inverting such matrices [27]. We also see that the process noise terms of $\mathbf{R}$ do not experience the correlation induced by the dynamics of the plant. Any correlation in the process noise arises solely from the moving average terms which distribute this noise.

3.3 Estimation Statistics

The parameters to be estimated are embedded in a stochastic system. Therefore, we are concerned about the statistics of the estimate. Namely, we want to know if the estimates are biased and we wish to predict the estimation error covariance. Under the assumptions of system-model linearity and Gaussian noise, the calculations required are straight-forward. However, even for a linear system, the parameter estimation problem is *nonlinear*, significantly complicating the statistics calculation problem. We can, however, linearize the problem in order to gain some insights about the error statistics. This linearization is accomplished by ignoring the random nature of the regressor matrix. Thus, the problem is reduced to a simple linear regression through which we can calculate an estimate of the statistics of the estimation errors. Note that this process whereby we ignore the random nature of the $\mathbf{R}$ matrix is designed to give us a feeling for the error statistics associated with the weighted least squares estimate. Since $\mathbf{R}$ is actually populated with ran-

dom quantities, the statistics calculations are actually much more complicated than presented here.

Consider a linear estimation problem incorporating the abbreviated notation:

$$
\begin{aligned}
\mathbf{y} &= \mathbf{H}\boldsymbol{\theta} \\
\mathbf{z} &= \mathbf{y} + \mathbf{n} \\
\mathcal{E}\{\mathbf{n}\} &= \mathbf{0} \\
\mathcal{E}\{\mathbf{n}\mathbf{n}'\} &= \mathbf{R}
\end{aligned}
\tag{72}
$$

with the minimum variance parameter vector estimate given by

$$
\hat{\boldsymbol{\theta}} = \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}'\mathbf{R}^{-1}\mathbf{z}
\tag{73}
$$

In order to determine the bias on the estimate, calculate the expected value of the estimation error, recognizing that $\mathbf{y}$ and $\boldsymbol{\theta}$ are deterministic quantities so $\mathcal{E}\{\boldsymbol{\theta}\} = \boldsymbol{\theta}$ and $\mathcal{E}\{\mathbf{y}\} = \mathbf{y}$:

$$
\begin{aligned}
\mathcal{E}\{\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\} &= \mathcal{E}\left\{\boldsymbol{\theta} - \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}'\mathbf{R}^{-1}\mathbf{z}\right\} \\
&= \mathcal{E}\left\{\boldsymbol{\theta} - \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}'\mathbf{R}^{-1}(\mathbf{y}+\mathbf{n})\right\} \\
&= \mathcal{E}\left\{\boldsymbol{\theta} - \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}'\mathbf{R}^{-1}\mathbf{y} - \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}'\mathbf{R}^{-1}\mathbf{n}\right\} \\
&= \boldsymbol{\theta} - \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}'\mathbf{R}^{-1}\mathbf{y} \\
&= \left\{\mathbf{I} - \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]\right\}\boldsymbol{\theta} \\
&= \{\mathbf{I} - \mathbf{I}\}\boldsymbol{\theta} = \mathbf{0}
\end{aligned}
\tag{74}
$$

Thus, in the linear case, the estimate is unbiased. Now, we can compute the estimation error variance. Let the estimation error be given by $e \doteq \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$. The estimation error covariance is computed by

$$
\begin{aligned}
\mathcal{E}\{\mathbf{e}\mathbf{e}'\} &= \mathcal{E}\left\{\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)'\right\} \\
&= \mathcal{E}\left\{\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}' - \hat{\boldsymbol{\theta}}\boldsymbol{\theta}' - \boldsymbol{\theta}\hat{\boldsymbol{\theta}}' + \boldsymbol{\theta}\boldsymbol{\theta}'\right\} \\
&= \mathcal{E}\left\{\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}'\right\} - \mathcal{E}\{\hat{\boldsymbol{\theta}}\}\boldsymbol{\theta}' - \boldsymbol{\theta}\mathcal{E}\{\hat{\boldsymbol{\theta}}'\} + \boldsymbol{\theta}\boldsymbol{\theta}' \\
&= \mathcal{E}\left\{\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}'\right\} - \boldsymbol{\theta}\boldsymbol{\theta}'
\end{aligned}
\tag{75}
$$

Now,

$$
\begin{aligned}
\mathcal{E}\{\hat{\theta}\hat{\theta}'\} &= \mathcal{E}\left\{ \left(\left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \mathbf{H}'\mathbf{R}^{-1}\mathbf{z} \right) \left(\mathbf{z}'\mathbf{R}^{-1}\mathbf{H} \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \right) \right\} \\
&= \mathcal{E}\left\{ \left(\left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \mathbf{H}'\mathbf{R}^{-1}(\mathbf{y}+\mathbf{v}) \right) \left((\mathbf{y}+\mathbf{v})'\mathbf{R}^{-1}\mathbf{H} \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \right) \right\} \\
&= \mathcal{E}\left\{ \left(\left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \mathbf{H}'\mathbf{R}^{-1}(\mathbf{H}\theta+\mathbf{v}) \right) \times \right. \\
&\qquad\qquad \left. \left((\mathbf{H}\theta+\mathbf{v})'\mathbf{R}^{-1}\mathbf{H} \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \right) \right\} \\
&= \mathcal{E}\left\{ \left(\theta + \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \mathbf{H}'\mathbf{R}^{-1}\mathbf{v} \right) \left(\theta' + \mathbf{v}'\mathbf{R}^{-1}\mathbf{H} \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \right) \right\} \\
&= \theta\theta' + \theta\mathcal{E}\left\{ \mathbf{v}'\mathbf{R}^{-1}\mathbf{H} \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \right\} + \mathcal{E}\left\{ \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \mathbf{H}'\mathbf{R}^{-1}\mathbf{v} \right\}\theta' + \\
&\qquad\qquad \mathcal{E}\left\{ \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \mathbf{H}'\mathbf{R}^{-1}\mathbf{v}\mathbf{v}'\mathbf{R}^{-1}\mathbf{H} \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \right\} \\
&= \theta\theta' + \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \mathbf{H}'\mathbf{R}^{-1}\mathcal{E}\{\mathbf{v}\mathbf{v}'\}\mathbf{R}^{-1}\mathbf{H} \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \\
&= \theta\theta' + \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \mathbf{H}'\mathbf{R}^{-1}\mathbf{H} \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \\
&= \theta\theta' + \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1}
\end{aligned}
\tag{76}
$$

Finally, we combine Equations (75) and (76) to get

$$
\mathcal{E}\{\mathbf{e}\mathbf{e}'\} = \left[\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\right]^{-1}
\tag{77}
$$

We must reiterate that the above calculations are rigorous if and only if the attendant estimation problem entails a linear regression, which it does not. Since we are estimating the parameters defining a dynamical system's model, the estimation problem in System ID is in fact nonlinear.

3.4 Implementation

Implementation of the batch system identification algorithm discussed in this chapter requires that we know the parameters which make up the equation error covariance matrix ($\mathbf{R}$). However, if we *know* the parameters, then identification is not required. Therefore, we propose an iterative scheme in which the parameter estimates are fed back into the linear minimum variance algorithm, obviating the need for knowledge of the parameter values.

41

The feedback is seen in the equation relating the current parameter estimate with previous estimates. After slight modification, Equation (49) is rewritten

$$\widehat{\boldsymbol{\theta}}_{k+1} = \left[\mathbf{H}_{k-1}^{(q)}{}'\mathbf{R}_k^{-1}\mathbf{H}_{k-1}^{(q)}\right]^{-1}\mathbf{H}_{k-1}^{(q)}{}'\mathbf{R}_k^{-1}\mathbf{z}_{k+1}^{(q)} \tag{78}$$

where $\mathbf{R}_k$ is a function of the previous parameter estimates. Clearly, $\mathbf{R}_k$ is also a function of time since new parameter estimates are generated with each time step.

$$\mathbf{R}_k = \mathbf{V}_k + \boldsymbol{\Gamma}_k \tag{79}$$

with $\mathbf{V}$ and $\boldsymbol{\Gamma}_k$ given by Equation (65) and Equation (66), respectively. However, the ν's are computed from the previous parameter estimate. This innovative idea, recursively calculating the weighting matrix, is used extensively throughout this research. Appendix A documents a concise algorithm which performs this recursion and generates parameter estimates. The measurement and process noise variances (σ_v^2 and σ_w^2) can now be treated as tuning parameters.

The algorithm implemented here assumes no process noise, which gives us the minimum variance weighting matrix as a symmetric, Toeplitz matrix (see Equation (65)) built from the estimated parameters with a scalar multiple. Since the Minimum-Variance Weighted Least Squares parameter estimate requires the product of this weighting matrix and its inverse, the scalar measurement noise variance parameter cancels. Thus, in order to obtain the parameter estimate, we require no knowledge of σ_v if we assume no process noise; this is a most pleasant aspect of the proposed System Identification approach.

An estimate of the estimation error variance can also be obtained. Equation (77) is modified to reflect the time- varying $\mathbf{R}$ matrix:

$$\widehat{\mathbf{R}_{\boldsymbol{\theta}}}_k = \left[\mathbf{H}'\mathbf{R}_k^{-1}\mathbf{H}\right]^{-1} \tag{80}$$

where $\widehat{\mathbf{R}_{\boldsymbol{\theta}}}_k$ is the estimate of $\mathcal{E}\{\mathbf{e}\mathbf{e}'\}$. The diagonal elements of $\widehat{\mathbf{R}_{\boldsymbol{\theta}}}_k$ provide us with estimates of the estimation error variances (within a scalar multiple) for each of the

parameters. These quantities could be combined to gauge the Excitation Level, as described in Chapter I.

Clearly, and despite of its linear appearance, this System Identification scheme is nonlinear. Thus, we cannot readily make broad statements about the stability of the algorithm. Also, the ability of the algorithm to capture the correct parameters when faced with poor initialization may be in question, as is always the case in nonlinear mathematics, where, e.g., convergence of iterative schemes is determined by a sufficiently close initial guess. Therefore, some experiments to motivate the use of the proposed ID algorithm are conducted and documented in the sequel.

3.5 Experimental Results

The importance of properly incorporating the weighting matrix in the ID cannot be seen readily from the previous discussion. Granted, we derived this Weighted Least Squares scheme specifically to reduce the parameter estimate covariance, viz., our estimate is a minimum- variance estimate; but how much effect do we see in practice? The answer is given by well-planned and careful experimentation. To this end, we present some experimental results which clearly demonstrate the benefits of proper weighting. In each of the experiments documented here, we excite the plant with white noise and corrupt the output measurements with another independent white noise sequence designed to produce a relatively low SNR of 20 dB. The Minimum-Variance Weighted Least Squares algorithm detailed in Appendix A is implemented in Matlab®. We employ this algorithm to generate parameter estimates for 1000 different batches of data, using the previously computed parameter estimates to build the weighting matrix. We also generate unweighted Least Squares estimates of the parameters from the same 1000 batches of data for comparison. The widely-used 'LS' command from the Matlab® System Identification Toolbox incorporates this unweighted Least Squares algorithm.

Table 1 Unweighted Least Squares estimate errors of the parameters defining a low-pass system with lightly damped poles.

Coefficient	Value	Mean Error	RMS Error	RMS Error (%)
a_1	-1.537e+00	4.847e-01	5.327e-01	3.466e+01%
a_2	9.025e-01	-4.295e-01	4.730e-01	5.241e+01%
b_1	1.923e-01	-6.864e-03	1.604e-01	8.341e+01%
b_2	1.731e-01	1.025e-01	1.902e-01	1.099e+02%

Table 2 Weighted Least Squares estimate errors of the parameters defining a low-pass system with lightly damped poles.

Coefficient	Value	Mean Error	RMS Error	RMS Error (%)
a_1	-1.537e+00	1.393e-02	4.216e-02	2.743e+00%
a_2	9.025e-01	-1.120e-02	3.900e-02	4.321e+00%
b_1	1.923e-01	-5.480e-03	8.260e-02	4.296e+01%
b_2	1.731e-01	1.855e-02	9.705e-02	5.608e+01%

3.5.1 Experiment 1. The first system we examine is described by the transfer function

$$\mathbf{G}(z) = \frac{b_1 z + b_2}{z^2 + a_1 z + a_2} = \frac{0.1923(z + 0.9)}{z^2 - 1.5371z + 0.9025} \tag{81}$$

which represents a low-pass system with a high-frequency zero and lightly damped poles:

$$p_1, p_2 = 0.95\angle \pm \frac{\pi}{5}$$

We present the results of the comparison in Tables 2 and 1. Clearly, the iteratively weighted System ID scheme produces significantly smaller estimation errors. Mean errors (bias) and RMS errors drop markedly when the weighting is used properly. However, we may still question the validity of either set of estimates. Also, the coefficients themselves are not what we typically concern ourselves with as control system designers. Rather, the frequency response and pole/zero locations are often more germane to control system design. Therefore, we also present graphical comparisons relating pole/zero locations and frequency domain errors.

(a) Unweighted Least Squares   (b) Weighted Least Squares
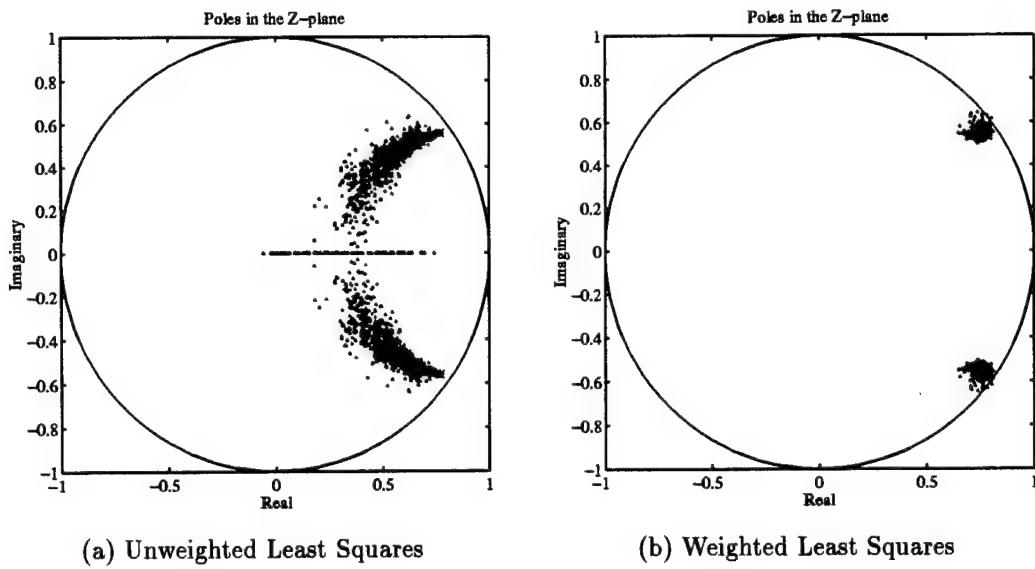
Figure 6.   Root Locations for Weighted and Unweighted Least Squares Estimates: *The plant is low-pass with lightly damped poles. (a) shows the root locations for all 1000 runs of the ID algorithm without proper weighting while (b) shows the pole locations for ID with weighting. The unit circle is superimposed for reference.*

First, Figures 6(a) and 6(b) show the root locations of the identified transfer functions' pole locations. Notice the significant spread of the roots in the Least Squares case relative to the tight grouping of poles in the weighted approach.

Root location notwithstanding, we are concerned with the identified transfer functions' frequency responses. Figure 7 gives a good indication of the increase in accuracy we get from weighting the estimates. Properly weighting the estimates results in much tighter worst-case envelopes and smaller RMS errors in both magnitude and phase Bode plots. Of particular interest are the low-frequency errors. Notice that the unweighted estimates exhibit significantly more error at low frequencies. This type of error is particularly distressing to the controls engineer, since most modern robust control system synthesis techniques require small plant uncertainty at lower frequencies and relegate the plant uncertainty to higher frequencies. Good

45
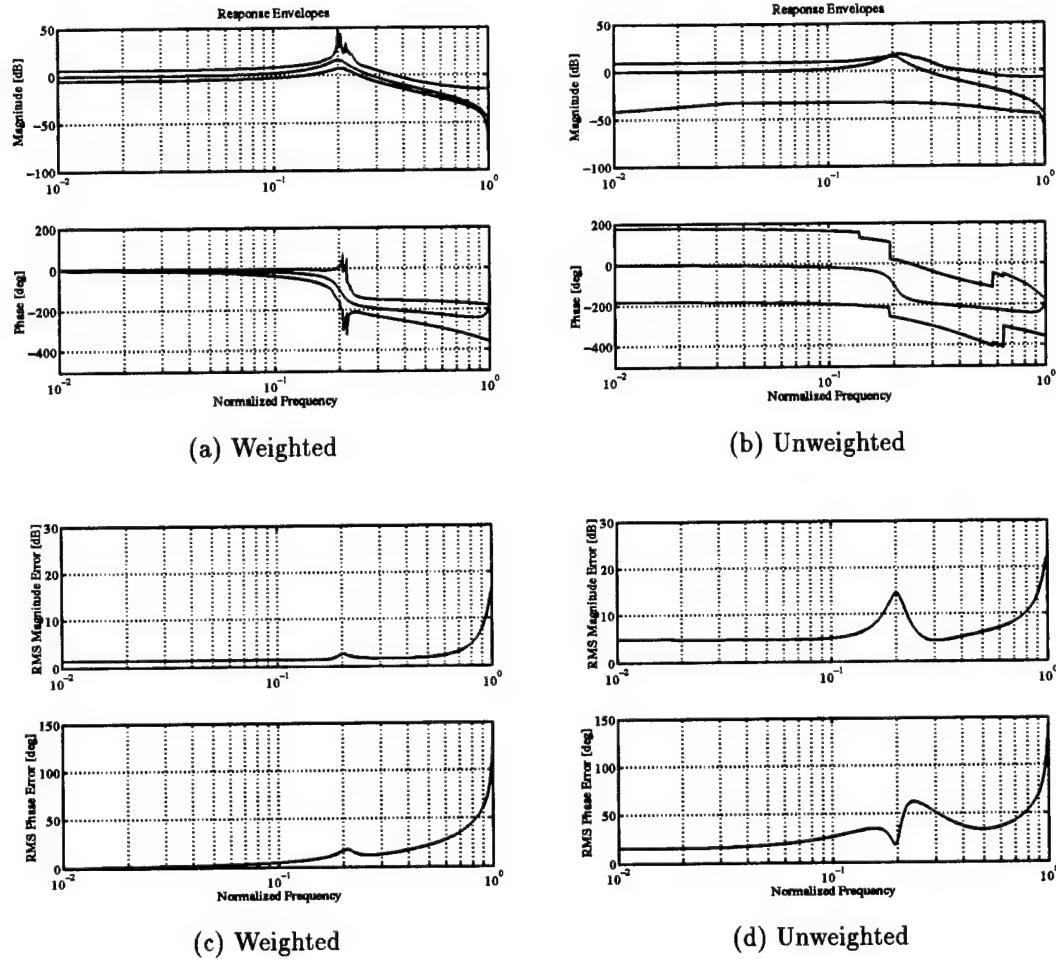
Figure 7. Frequency Response Envelopes and Errors: *The plant is low-pass with a pair of lightly damped poles; (a) shows the worst-case envelopes for the magnitude and phase frequency response curves when the transfer functions are identified via Weighted Least Squares. (b) gives these curves without the benefit of weighting. Note that the envelope plots include the nominal, largest, and smallest estimates for each frequency. (c) and (d) show the RMS errors in the responses for the weighted and unweighted identification, respectively.*

discussions of these error-distribution (i.e. uncertainty distribution) requirements for robust control design are given by Doyle [17] and Ridgely [74].

Table 3 Weighted Least Squares estimate errors of the parameters defining a low-pass system with all real poles.

Coefficient	Value	Mean Error	RMS Error	RMS Error (%)
a_1	-2.100e+00	1.088e-01	1.557e-01	7.414e+00%
a_2	1.460e+00	-1.923e-01	2.796e-01	1.915e+01%
a_3	-3.360e-01	8.663e-02	1.288e-01	3.833e+01%
b_1	2.400e-02	3.471e-03	4.944e-03	2.060e+01%

Table 4 Unweighted Least Squares estimate errors of the parameters defining a low-pass system with all real poles.

Coefficient	Value	Mean Error	RMS Error	RMS Error (%)
a_1	-2.100e+00	1.366e+00	1.374e+00	6.543e+01%
a_2	1.460e+00	-1.665e+00	1.674e+00	1.147e+02%
a_3	-3.360e-01	3.794e-01	3.980e-01	1.185e+02%
b_1	2.400e-02	-1.815e-03	1.724e-02	7.184e+01%

3.5.2 Experiment 2.   Our next experiment involves identification of a plant with all real poles. The transfer function is given by

$$\mathbf{G}(z) = \frac{0.0024z}{(z - 0.6)(z - 0.7)(z - 0.8)} \tag{82}$$

We present the identification results for comparison in Tables 3 and 4. Once again, the Weighted Least Squares ID scheme yields significantly more accurate estimates of the parameters.

However, the increase in accuracy is more striking in the root locations and the frequency response. Figure 8 shows the large amount of spread in the pole locations when ID is performed with no weighting. In particular, note that the poles are shifted into much higher frequencies, suggesting significant errors in the frequency response. Figure 9 confirms these suspicions; the worst-case envelopes and the RMS errors are significantly improved by weighting the estimates. In particular, we see that the weighted ID yields transfer functions with better-behaved low-frequency errors. The Weighted Least Squares ID produces low-frequency RMS errors around 1 dB, while the unweighted estimates exhibit very large low-frequency errors of about 20 dB.

47

(a) Unweighted Least Squares   (b) Weighted Least Squares
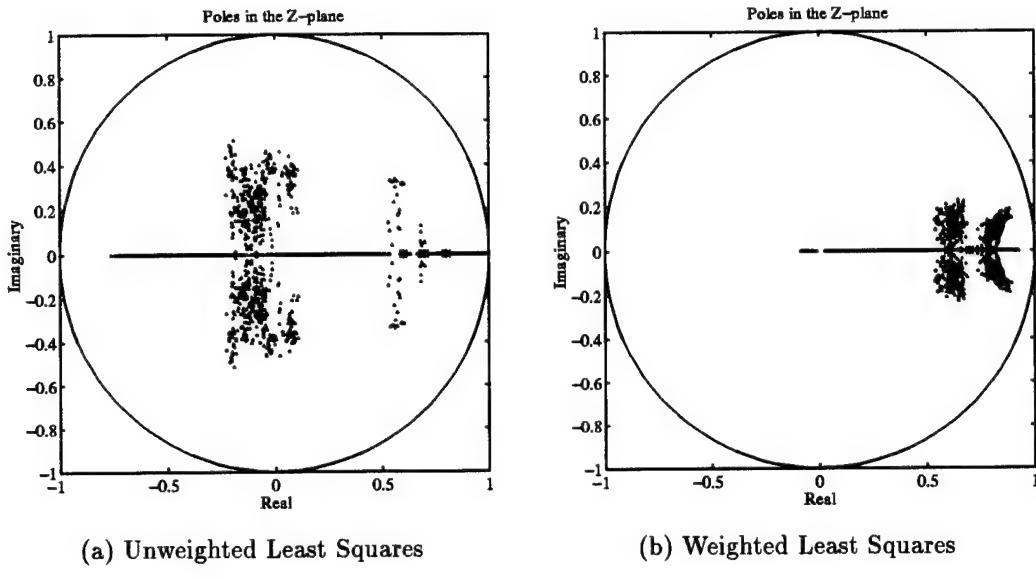
Figure 8.   Root Locations for Weighted and Unweighted Least Squares Estimates:
*The plant is low-pass with all real poles. (a) shows the root locations for
all 1000 runs of the ID algorithm without proper weighting while (b) shows
the pole locations for ID with weighting. The unit circle is superimposed
for reference.*

Thus, the transfer function estimates derived by standard Least Squares estimation
are practically unusable as models for control system synthesis.

3.5.3   Experiment 3.   As a final experiment, we consider a plant with lightly
damped poles and zeros. The truth model transfer function is given by

$$\mathbf{G}(z) = \frac{b_1 z^2 + b_2 z + b_3}{z^4 + a_1 z^3 + a_2 z^2 + a_3 z + a_4}$$
$$= \frac{0.2078 z^2 - 0.3295 z + 0.19957}{z^4 - 3.0161 z^3 + 4.0683 z^2 - 2.8967 z + 0.92237}$$

which represents a system with poles

$$p_1, p_2 = 0.98 \angle \pm \tfrac{\pi}{10}$$
$$p_3, p_4 = 0.98 \angle \pm \tfrac{3\pi}{10}$$

and zeros

$$\zeta_1, \zeta_2 = 0.98 \angle \pm \tfrac{\pi}{5}$$

48
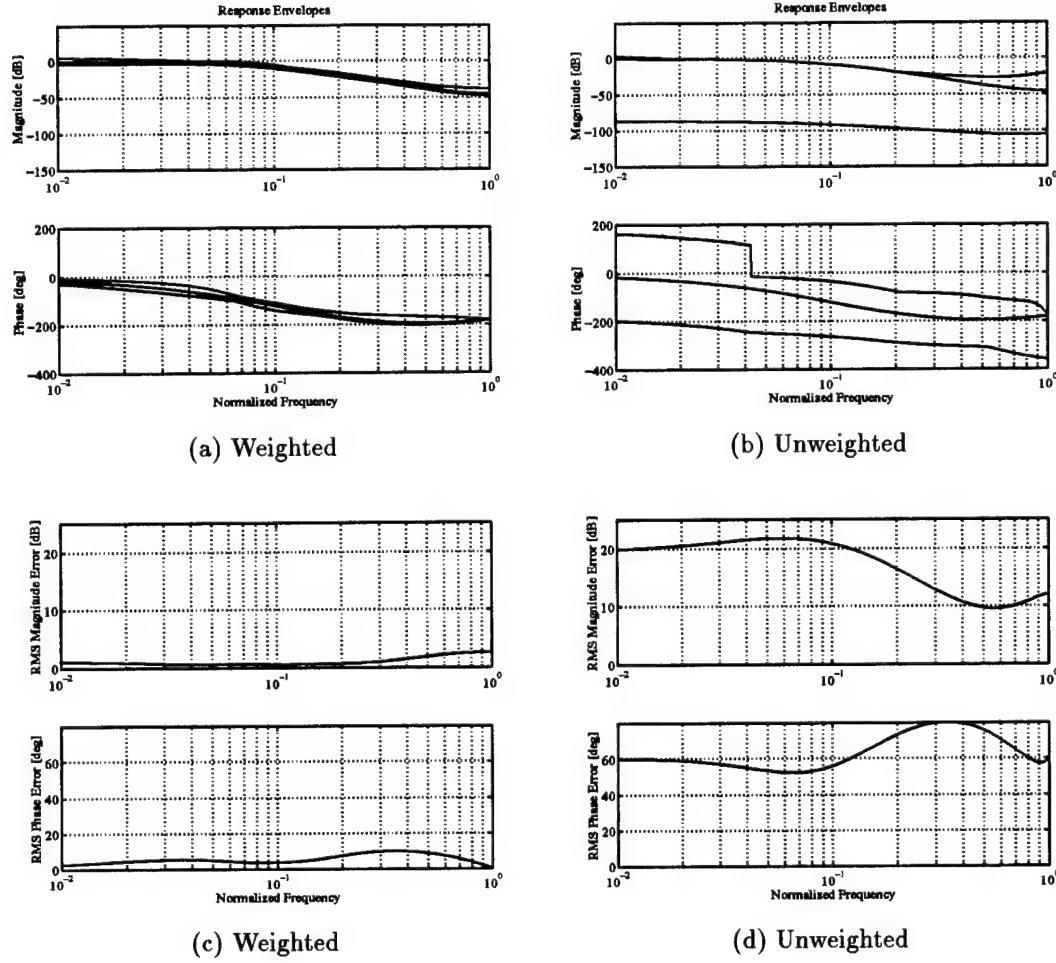
Figure 9. Frequency Response Envelopes and Errors: *The plant is low-pass with all real poles; (a) shows the worst-case envelopes for the magnitude and phase frequency response curves when the transfer functions are identified via Weighted Least Squares. (b) gives these curves without the benefit of weighting. Note that the envelope plots include the nominal, largest, and smallest estimates for each frequency. (c) and (d) show the RMS errors in the responses for the weighted and unweighted identification, respectively.*

This system is designed specifically to create a complex shape in the frequency domain and to illustrate the effect weighting has on the identified zeros. Tables 5 and 6 document the accuracy of the parameter estimates for the weighted and unweighted least squares estimates, respectively. The pattern established in previous experiments is continued here; the properly weighted estimates are markedly

49

Table 5 Weighted Least Squares estimate errors of the parameters defining a low-pass system with lightly damped poles and zeros.

| Coefficient	Value	Mean Error	RMS Error	RMS Error (%)
a_1	-3.016e+00	6.508e-03	1.542e-02	5.112e-01%
a_2	4.068e+00	-2.005e-02	4.125e-02	1.014e+00%
a_3	-2.897e+00	2.277e-02	4.523e-02	1.562e+00%
a_4	9.224e-01	-9.514e-03	1.859e-02	2.016e+00%
b_1	2.078e-01	3.138e-03	2.237e-02	1.077e+01%
b_2	-3.295e-01	-8.006e-03	4.251e-02	1.290e+01%
b_3	1.996e-01	5.748e-03	2.436e-02	1.221e+01%

Table 6 Unweighted Least Squares estimate errors of the parameters defining a low-pass system with lightly damped poles and zeros.

| Coefficient	Value	Mean Error	RMS Error	RMS Error (%)
a_1	-3.016e+00	2.079e+00	2.089e+00	6.925e+01%
a_2	4.068e+00	-3.898e+00	3.904e+00	9.597e+01%
a_3	-2.897e+00	3.135e+00	3.139e+00	1.084e+02%
a_4	9.224e-01	-1.070e+00	1.078e+00	1.169e+02%
b_1	2.078e-01	-9.171e-03	8.538e-02	4.109e+01%
b_2	-3.295e-01	4.395e-01	4.461e-01	1.354e+02%
b_3	1.996e-01	-2.006e-01	2.104e-01	1.054e+02%

improved over the unweighted estimates. Furthermore, the accuracy of the root locations (both poles and zeros) is greatly improved by weighting the estimates, as we see in Figure 10. We should expect to see a large improvement in the frequency response errors by weighting the estimates. Figure 11 confirms this expectation. The transfer functions' frequency response curves derived via unweighted parameter estimates hardly resemble the underlying plant's. In fact, Figure 11(b) shows that the envelope created by the largest and smallest estimated frequency responses does not even include the plant's response curve for some frequencies. In contrast, the frequency response envelope based on Minimum Variance/Weighted estimates is very tight. The RMS error curves further confirm the superiority of our Minimum Variance/Weighted Least Squares ID scheme.

(a) Weighted Zero Locations   (b) Unweighted Zero Locations

(c) Weighted Pole Locations   (d) Unweighted Pole Locations
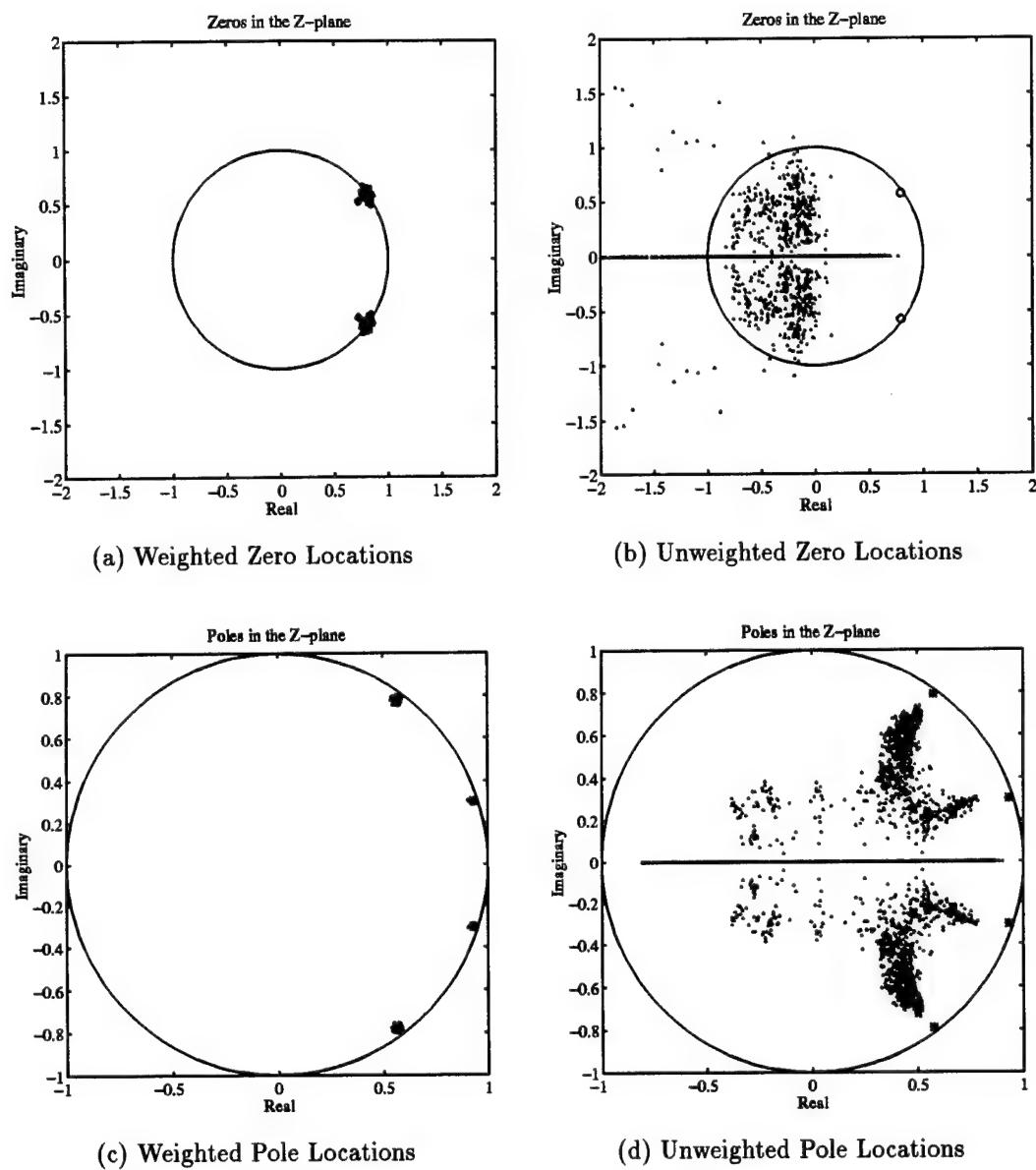
Figure 10.   Weighted and Unweighted Pole/Zero Locations: *The plant has four lightly damped poles and two zeros very close to the unit circle. (a) and (b) show the weighted and unweighted estimated zero locations. The estimated pole locations are shown in (c) and (d).*

3.6   Conclusions

In this chapter, we saw how the equation error in linear regression is affected by the dynamics of the system under test, resulting in a weighting matrix which

51
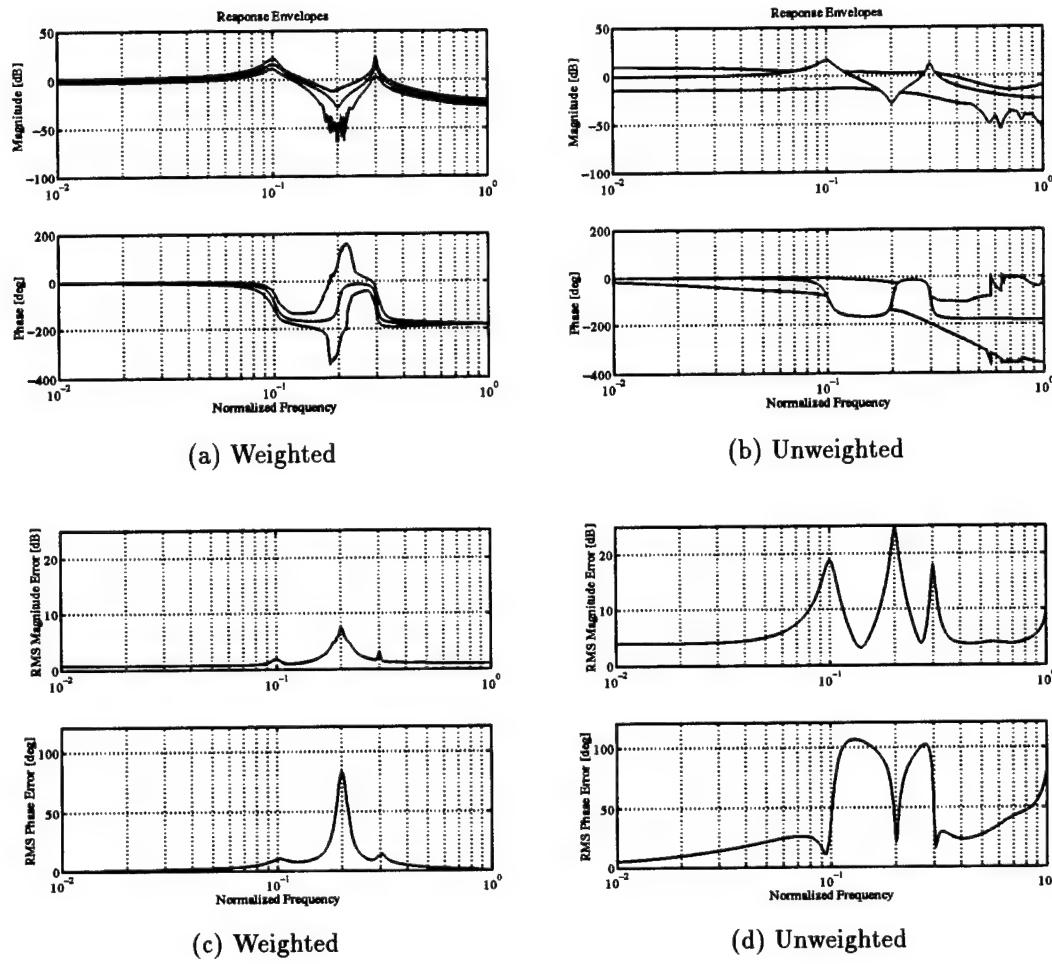
Figure 11.   Frequency Response Envelopes and Errors: *The plant has four lightly damped poles and two zeros very close to the unit circle. (a) shows the worst-case envelopes for the magnitude and phase frequency response curves when the transfer functions are identified via Weighted Least Squares. (b) gives these curves without the benefit of weighting. Note that the envelope plots include the nominal, largest, and smallest estimates for each frequency. (c) and (d) show the RMS errors in the responses for the weighted and unweighted identification, respectively.*

is a function of the parameters under scrutiny. Although we do not have knowledge of the actual parameters defining the system, we proposed and experimentally tested a method by which the current estimates of the parameters are used iteratively to construct this matrix. This also allows the algorithm to adapt to slowly changing/drifting parameters. Furthermore, in our analysis, we derived a candidate

52

estimate for the estimation error variances which may be used to gauge the Excitation Level. The experiments clearly show the advantage of properly weighting the parameter estimates. In fact, for the three plants used in this study, the widely-used (unweighted) Least Squares method was incapable of producing usable plant models when the SNR was as low as 20 dB. In contrast, the Minimum Variance/iteratively weighted algorithm developed here and outlined in Appendix A yielded plant models which closely reproduce the true plant in both root locations and frequency response.

Having developed and tested an efficient ID algorithm, we are now faced with the question, "How can ID be enhanced by the inputs?" It stands to reason that an identification scheme's accuracy is limited by the quality of the raw data available to it. Therefore, since ID uses measurements of the inputs and outputs of the plant, we want these quantities to encode as much information as possible about the unknown parameters. In Chapter IV, we address this idea of 'information' and derive input power spectral density functions which maximize the parameter information contained in the input and output signals.

IV. Information Theory and Optimal Input Shaping Filters

4.1 Introduction

System Identification (ID) consists of deriving, from knowledge of input signals applied to an unknown plant and measured, noise-corrupted plant output signals, an estimate of a vector of parameters which is used to form a mathematical model describing the behavior of the plant. The "plant" in question is a dynamical system. For example, the parameters might be the coefficients of a linear differential or difference equation. Alternatively, the ID algorithm may directly seek the locations of poles and zeros of a transfer function which describes a control system. A vast amount of work has been accomplished with an aim towards finding an optimal solution to the very basic problem of identifying the parameters of dynamical systems. For example, Åström and Eykhoff [4] present a good survey of the state-of-the-art (ca. 1971) in ID. A more current and widely-used reference is [42]. Scores of other sources exist which document the level of energy devoted to the field of System ID. Thus, the problem of identifying the unknown parameters of a dynamical system, though not completely solved, is a well-established field of research. Unfortunately, good solutions to the System ID problem have yet to be devised. Our work strives to contribute toward a practical solution of the System ID problem.

However the solution to the ID problem is approached, the analyst often must accept the input/output data as it arrives. However, in some cases, the analyst might exercise control over the input signal. In these cases, some control over the excitation applied to the unknown plant should help enhance the performance of the ID algorithm, especially in the face of measurement and process noise. Surely, this idea is not original, as is evidenced by the literature. For example, Mehra [55, 56] presents several approaches for achieving optimal excitation, and Olmstead, in his Ph.D. dissertation [67], addresses the problem of ID within a feedback control loop. A common theme exists throughout the literature. Namely, the solution is closely

linked to the type of ID to be performed; thus, the input is designed to limit, or optimize (minimize), some metric applied to the estimation error of the ID algorithm.

A new approach to the optimal input design problem for ID is proposed in this chapter. The concepts of information theory (traditionally a communications engineer's tool) and stochastic processes are exploited to maximize the 'information' content of the unknown plant's output. In this dissertation, the System ID process is being cast into a communications framework as follows. The message to be transmitted is the plant's data, i.e. the parameters defining the model of the plant. This message is encoded into the "transmitted" signal which consists of the plant's input and output measurements, the latter corrupted by measurement noise in the "communications channel." If the 'information' is properly encoded at the transmitter's end such that the signals under scrutiny by the ID algorithm at the receiver's end contain as much knowledge as possible about the unknown parameters, then the potential for successful ID will be enhanced.

Thus, we present this approach to optimal inputs via maximizing 'information' in the following manner. First, in Section 4.2 we offer an overview of the pertinent information theory, concentrating on one particular definition of the term 'information.' Next, in Section 4.3 we address specific applications of this definition to stochastic processes, leading to optimal shaping filter design. We outline different approaches to the design of the optimal shaping filter, both for a deterministic plant and an unknown (or partially known) plant. Finally, we offer some conclusions in Section 4.6.

4.2   Information Theory

The term 'information' can mean many things to many people. In fact, as Cole [13] points out, 'information' can take on two dichotomous meanings. In some contexts, information is taken as the reduction of uncertainty about the received message. Certainly, this definition is most intuitive if the receivers point of view

55

is taken. The receiver wishes to extract, with the least uncertainty, the message embedded within some (noise-corrupted) signal. However the case can also be made that information can be defined as an "increase in uncertainty." This latter definition makes most sense from a sender's perspective. If the sender of a message draws from a very rich set of messages to send, then each message sent contains a large amount of information. Conversely, if a sender is limited to a very small set of messages, then each message sent contains a small amount of information. Thus, consider the degenerate case in which the set of messages is a singleton; the sender can send only one message. Clearly, no information is imparted to the receiver, since the receiver knows the message with no ambiguity without even monitoring the signal!

We are using concepts which are usually applied to communications theory; our aim is to apply information theoretic concepts to System Identification. A short glossary of information/communication theory nomenclature is included here in order to establish concisely the analogy between ID and communications.

message The relevant facts encoded in the **signal**. In the System ID experiment design context, the **message** consists of the vector of plant parameters (θ).

sender The *designer* of the encoding scheme. The **sender** *does not* know the **signal** sent. Rather, he knows some statistics about the **signal** to be sent and about most quantities involved, e.g., the communications channel. In our work, the **sender** is the input generator.

signal The sequence of numbers actually arriving at the **receiver**. Our signal consists of inputs to the plant and measured plant outputs. Only the latter is sent through the communications channel, and is, therefore, noise corrupted. (In some applications, however, the input is also noise-corrupted.)

telegraphist The process by which the **signal** is sent. In our context, the **telegraphist** is the plant. We note that the **telegraphist** 'knows' the **message**.

receiver The entity that receives the **signal**, and attempts to decode the **message**. Our **receiver** is the ID algorithm.

Our goal is to provide the ID algorithm (e.g. our Minimum Variance/Weighted Least Squares algorithm) with as much information as possible concerning the unknown plant. Thus, the optimal input generator takes on the role of the sender of the message, motivating the latter definition for 'information.' We need now a rigorous mathematical definition for the terms 'information,' or 'uncertainty.' The following discussion, largely adopted from Papoulis' text [68] gives those required definitions.

4.2.1 Entropy as a Measure of Information. The previous discussion motivates and justifies the use of uncertainty as information for our application. Shannon's classical work on communication [80] provides the definition for uncertainty (information) accepted throughout the communications field. Shannon's work defines uncertainty as *entropy*, a quantity applied to partitions of sets. As interpreted by Papoulis [68], the entropy associated with a partition ($\mathcal{H}(\mathfrak{A})$ of the set $\mathfrak{A}$ with partitions $\mathcal{A}_i$) is derived such that the following three postulates are satisfied:

1. $\mathcal{H}(\mathfrak{A})$ is a continuous function of $p_i = P(\mathcal{A}_i)$ (the probabilities assigned to each event in the partition).

2. If $p_1 = p_2 = \cdots = p_N = \frac{1}{N}$, then $\mathcal{H}(\mathfrak{A})$ is an increasing function of N.

3. If a new partition, $\mathfrak{B}$, is formed by subdividing one of the sets of $\mathfrak{A}$, then $\mathcal{H}(\mathfrak{A}) \geq \mathcal{H}(\mathfrak{B})$.

Shannon shows that the functional, $\mathcal{H}$, is uniquely given by

$$\mathcal{H}(\mathfrak{A}) = -K \sum_{i=1}^{N} p_i \ln(p_i) \tag{83}$$

57

where $K > 0$ is arbitrary and may be used to establish the units of measure.

Setting $K = 1$ and adopting a subtle change of notation, Papoulis [68] defines entropy for a discrete random variable as follows:

Definition 4.2.1 (Entropy of a discrete random variable) *The* entropy *associated with a* discrete-valued *random variable* x, *denoted* $\mathcal{H}(\mathsf{x})$, *is given by*

$$\mathcal{H}(\mathsf{x}) = -\sum_{i=1}^{N} p_i \ln(p_i)$$

where p_i is the probability associated with each of the N possible realizations of x.

Extending the notion of entropy of a discrete-valued random variable to a random variable with a continuous distribution is accomplished through limiting the summation in such as way as to achieve an integral in place of the summation. Again, the derivation is covered in Papoulis [68] and yields the following definition:

Definition 4.2.2 (Entropy of a continuous random variable) *The* entropy *of a* continuous random variable, x, *denoted* $\mathcal{H}(\mathsf{x})$, *is given by*

$$\mathcal{H}(\mathsf{x}) = -\int_{-\infty}^{\infty} p_\mathsf{x}(x) \ln(p_\mathsf{x}(x)) dx$$

where $p_\mathsf{x}(x)$ is the probability density function associated with the continuous random variable x. *Also note the integral is taken over the region where $p_\mathsf{x}(x) \neq 0$.*

Careful inspection of the previous definitions allows us to represent the entropy of either a continuous or discrete random variable via the use of the expectation operator:

Definition 4.2.3 (Entropy as an expectation) *The* entropy *of a random variable,* x, *is given by*

$$\mathcal{H}(\mathsf{x}) = -\mathcal{E}_\mathsf{x}\{\ln(p_\mathsf{x}(\mathsf{x}))\}$$

where p_x is the probability density function associated with x *and $\mathcal{E}\{\cdot\}$ denotes the expectation operator.*

Entropy as defined addresses the uncertainty in one random variable, but we are typically concerned with many quantities, usually arranged in a vector. Thus, we derive and define the entropy in a random vector (joint entropy) and the entropy in a random variable (or vector) conditioned on some other quantity (conditional entropy). Joint entropy is given by a simple extension of Definition 4.2.3:

Definition 4.2.4 (Joint entropy) *The* joint entropy *of a random vector,* $\mathbf{x}$*, is given by*

$$\mathcal{H}(\mathbf{x}) = -\mathcal{E}_{\mathbf{x}}\{\ln(f(\mathbf{x}))\}$$

where $f(\mathbf{x})$ *is the probability density function associated with* $\mathbf{x}$*.*

Finally, conditional entropy is the uncertainty of a random quantity given some knowledge about some other quantity. Since the quantities in question may be either scalars or vectors, we present the following definitions in vector notation, which may be specialized to the scalar case.

Definition 4.2.5 (Conditional Entropy) *The* conditional entropy *of a random vector,* $\mathbf{x}$ *given that a particular realization of* $\mathbf{y}$ *has occurred, denoted by* $\mathcal{H}(\mathbf{x}|\mathbf{y})$*, is given by*

$$\begin{aligned}
\mathcal{H}(\mathbf{x}|\mathbf{y}) &= -\mathcal{E}_{\mathbf{x}}\{\ln(f(\mathbf{x}|\mathbf{y})|\mathbf{y} = \mathbf{y}\} \\
&= -\int f(\mathbf{x}|\mathbf{y})\ln(f(\mathbf{x}|\mathbf{y}))d\mathbf{x}
\end{aligned}$$

Taking all possible realizations of $\mathbf{y}$ *into account, the* conditional entropy *of* $\mathbf{x}$ *assuming some unknown realization of* $\mathbf{y}$ *is given by*

$$\begin{aligned}
\mathcal{H}(\mathbf{x}|\mathbf{y}) &= \mathcal{E}_{\mathbf{y}}\{\mathcal{H}(\mathbf{x}|\mathbf{y})\} \\
&= \int f(\mathbf{y})\mathcal{H}(\mathbf{x}|\mathbf{y})d\mathbf{y} \\
&= -\iint f(\mathbf{x}, \mathbf{y})\ln(f(\mathbf{x}|\mathbf{y}))d\mathbf{x}d\mathbf{y}
\end{aligned}$$

4.2.2 Mutual Information. Entropy, taken as information about a quantity of interest, may not be a meaningful quantity in and of itself. Rather, we often desire a quantification of the improvement in information resulting from observations of related quantities. For example, this work addresses increasing the information about a set of unknown parameters contained in the inputs and outputs taken from a plant under test. This increase of information about one random variable contained in another variable leads us to the concept of mutual information, defined below:

Definition 4.2.6 (Mutual Information) *The* mutual information *between two random vectors, denoted by $I(\mathbf{x}, \mathbf{y})$, is given by*

$$\mathcal{I}(\mathbf{x}, \mathbf{y}) = \mathcal{H}(\mathbf{x}) + \mathcal{H}(\mathbf{y}) - \mathcal{H}(\mathbf{x}, \mathbf{y})$$

Applying the definitions for entropy and joint entropy, we see

$$\mathcal{I}(\mathbf{x}, \mathbf{y}) = \mathcal{E}\left\{\ln \frac{f(\mathbf{x}, \mathbf{y})}{f(\mathbf{x})f(\mathbf{y})}\right\} \tag{84}$$

where $f(\mathbf{x}, \mathbf{y})$ is the joint probability density function of $\mathbf{x}$ and $\mathbf{y}$. Since $f(x, y) = f(x|y)f(y)$ and using Definition 4.2.5, we have

$$\begin{aligned} \mathcal{I}(\mathbf{x}, \mathbf{y}) &= \mathcal{H}(\mathbf{x}) - \mathcal{H}(\mathbf{x}|\mathbf{y}) \\ &= \mathcal{H}(\mathbf{y}) - \mathcal{H}(\mathbf{y}|\mathbf{x}) \end{aligned} \tag{85}$$

Equation (85) shows how mutual information is construed as the information contained in one quantity about another. Let us assume we are concerned with information about $\mathbf{y}$ contained in $\mathbf{x}$; then the second form of Equation (85) yields the amount of uncertainty in $\mathbf{y}$ less the uncertainty of $\mathbf{y}$ given that $\mathbf{x}$ has been observed. In other words, mutual information quantifies the reduction of uncertainty in $\mathbf{y}$ by observing $\mathbf{x}$. Thus, if we can maximize $\mathcal{I}(\mathbf{x}, \mathbf{y})$ by some choice of the statistics of $\mathbf{x}$, that particular distribution for $\mathbf{x}$ is optimal in that it yields maximal information about the unknown variable, $\mathbf{y}$.

4.3 Stochastic Processes – Entropy Rate

While information theory (specifically mutual information) can be exploited to help us determine the optimal inputs for ID, we are still faced with a possible 'curse of dimensionality.' ID schemes use a history of data, not just the instantaneous measurements of the current input and output. Thus, the sequences of input and output which appear in the equations describing entropy and mutual information would grow into very long vectors; in fact, steady-state analysis would require infinitely long vectors.

A useful tool for reducing the data requirements is the concept of *Entropy Rate*, defined in the sequel. Entropy rate quantifies the average information per sample in a block of consecutive samples taken from a process.

We can also reduce the data requirements through a transformation into the frequency domain. Our random vectors consist of input and output data which are interpreted as realizations of stochastic processes. Since our data are inherently discrete-time domain sequences, a natural candidate for transformation is the discrete-time Fourier transform (DTFT). Indeed, we shall see that the DTFT, denoted $\mathcal{F}\{\cdot\}$, is a valuable tool to be used in the ensuing optimization (maximization) of the information contained in the measured signals.

The statistics for a discrete-time random process are represented in terms of the joint probability density function $f(x_1, x_2, \ldots, x_m)$ for the random variables x_1, $\mathsf{x}_2, \ldots, \mathsf{x}_m$. The joint entropy for these m random variables

$$\mathcal{H}(\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_m) = -\mathcal{E}\{\ln(f(\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_m))\} \tag{86}$$

is known as the mth-order entropy of the discrete-time process x_k, where $\mathsf{x}_k \triangleq \mathsf{x}(t_k)$.

We will assume that all processes are strict sense stationary (i.e. all statistics are independent of the choice of time origin). With stationarity in place, the mth-order entropy quantifies the uncertainty associated with *any* consecutive m samples from the process.

Consider a Markoff process, x_k. Since this process is Markoff, we can write the joint probability density function for m consecutive variables in the sequence as a product of first order conditional densities

$$f(x_1, x_2, \ldots, x_m) = f(x_m|x_{m-1}) \cdots f(x_2|x_1)f(x_1) \tag{87}$$

Applying the definitions for joint and conditional entropies (Definitions 4.2.4 and 4.2.5) yields

$$\mathcal{H}(x_1, \ldots, x_m) = \mathcal{H}(x_m|x_{m-1}) + \cdots + \mathcal{H}(x_2|x_1) + \mathcal{H}(x) \tag{88}$$

where $\mathcal{H}(x)$ is the *first order entropy* for the process:

$$\mathcal{H}(x) = \mathcal{H}(x_k) \ \forall \ k \in \mathbb{Z}$$

Invoking stationarity, we have

$$\mathcal{H}(x_1, \ldots, x_m) = (m - 1)\mathcal{H}(x_1, x_2) - (m - 2)\mathcal{H}(x) \tag{89}$$

for the Markoff process x.

The next important concept involves the uncertainty about the present realization of a process when a portion of its past has been observed.

Definition 4.3.1 (Conditional Entropy of Order m) *The* Conditional Entropy *of Order m for a process x_k is given by the entropy of the current state of the process under the assumption that the m most recent values have been observed. For example,*

$$\mathcal{H}(x_n|x_{n-1}, \ldots, x_{n-m})$$

represents the conditional entropy of order m taken at time n.

Now, since obtaining more information about the past of a process can only decrease the uncertainty about the current value of the process, we conclude that the mth order condition entropy is a decreasing function in m [68]:

$$\mathcal{H}(x_n|x_{n-1}, \ldots, x_{n-m}) \leq \mathcal{H}(x_n|x_{n-1}, \ldots, x_{n-(m-1)}) \tag{90}$$

Now, if the mth order conditional entropy is bounded below, it will tend to some limit, motivating the following definition:

Definition 4.3.2 (Conditional Entropy of a Process) *The* Conditional Entropy *of a process* $\mathbf{x}_k$, *denoted by* $\mathcal{H}_c(\mathbf{x})$ *is given by*

$$\mathcal{H}_c(\mathbf{x}) = \lim_{m \to \infty} \mathcal{H}(\mathbf{x}_n | \mathbf{x}_{n-1}, \ldots, \mathbf{x}_{n-m})$$

and quantifies the measure of uncertainty about the present of the process assuming the entire past of the process has been observed.

As a special case, consider a Markoff process, **x**. Here,

$$\mathcal{H}(\mathbf{x}_n | \mathbf{x}_{n-1}, \ldots, \mathbf{x}_{n-m}) = \mathcal{H}(\mathbf{x}_n | \mathbf{x}_{n-1}) \tag{91}$$

Since the process is stationary, we can arbitrarily set the time indices, yielding

$$\mathcal{H}_c(\mathbf{x}) = \mathcal{H}(\mathbf{x}_1, \mathbf{x}_2) - \mathcal{H}(\mathbf{x}) \tag{92}$$

As we noted previously, we may be concerned with the average uncertainty per sample in a block of consecutive samples from a process. By considering this 'rate of entropy', we can quantify the information contained in an entire process history. Otherwise, the quantities involved would have to grow infinitely long. Thus, we make the following definition:

Definition 4.3.3 (Entropy Rate) Entropy Rate *of a process, denoted by* $\overline{\mathcal{H}}(\mathbf{x})$, *is the average information per sample for the entire history of the process:*

$$\overline{\mathcal{H}}(\mathbf{x}) = \lim_{m \to \infty} \frac{1}{m} \mathcal{H}(\mathbf{x}_1, \ldots, \mathbf{x}_m)$$

An important special case occurs when the process is Markoff. In this case,

$$\overline{\mathcal{H}}(\mathbf{x}) = \mathcal{H}(\mathbf{x}_1, \mathbf{x}_2) - \mathcal{H}(\mathbf{x}) = \mathcal{H}_c(\mathbf{x}) \tag{93}$$

which follows from Equation (89) and the definition for entropy rate. Thus, we see that entropy rate for a Markoff process converges to the conditional rate. The next theorem shows that this limit holds in general.

Theorem 4.3.1 *The entropy rate of a process,* $\mathbf{x}_n$, *equals the process's conditional entropy:*

$$\overline{\mathcal{H}}(\mathbf{x}) = \mathcal{H}_c(\mathbf{x})$$

Proof: Given a convergent sequence, a_k, then the following holds:

$$a_k \to a \Rightarrow \frac{1}{m} \sum_{k=1}^{m} a_k \to a \tag{94}$$

Now, $\mathbf{x}_n$ is stationary, so (as in Equation (88)), the joint entropy for the process is expressed as

$$\mathcal{H}(\mathbf{x}_1, \ldots, \mathbf{x}_m) = \mathcal{H}(\mathbf{x}) + \sum_{k=1}^{m} \mathcal{H}(\mathbf{x}_n | \mathbf{x}_{n-1}, \ldots, \mathbf{x}_{n-k}) \tag{95}$$

Thus,

$$\begin{aligned}
\overline{\mathcal{H}}(\mathbf{x}) &= \lim_{m \to \infty} \frac{1}{m} \mathcal{H}(\mathbf{x}_1, \ldots, \mathbf{x}_m) \\
&= \lim_{m \to \infty} \frac{1}{m} \mathcal{H}(\mathbf{x}) + \lim_{m \to \infty} \frac{1}{m} \left(\sum_{k=1}^{m} \mathcal{H}(\mathbf{x}_n | \mathbf{x}_{n-1}, \ldots, \mathbf{x}_{n-k}) \right)
\end{aligned}$$

Now, using the Definition 4.2.5 and Equation (94), we have the desired result. [68]

∎

As often is the case, the controls engineer is best served by transforming quantities of interest into the frequency domain. To this end, we derive entropy rate as a function of the power spectral density (PSD) of a signal of interest, viz., the entropy rate for a normal process is given by the following theorem.

Theorem 4.3.2 *Given a normal process,* $\mathbf{x}$, *with the associated PSD function denoted* $\mathcal{S}_{\mathbf{xx}}(\omega)$, *the entropy rate associated with* $\mathbf{x}$ *is given by*

$$\overline{\mathcal{H}}(\mathbf{x}) = \ln \sqrt{2\pi e} + \frac{1}{4\pi} \int_{-\pi}^{\pi} \mathcal{S}_{\mathbf{xx}}(\omega) d\omega$$

Proof: See Papoulis [68]. ∎

64

4.4 Discussion

Up to this point, we have given a condensed presentation of pertinent information theory, with specific emphasis on the quantities of interest to this research. Namely, we now have a workable definition for information, and more important, we have a way to compute the information rate contained in a stochastic process. Furthermore, the computation of the information rate involves frequency domain quantities, allowing the control engineer the freedom to work within that mathematically convenient framework. These tools are used in the sequel to derive the optimal PSD of the input to a linear plant; optimal in the sense that information about unknown parameters is maximized. Obviously, the said parameters eventually will be estimated by applying our specific System ID algorithm to the input and noise-corrupted outputs.

4.5 Optimal Shaping Filters

The previous sections provide us with a set of tools with which we can evaluate the information content of a signal. Thus, System Identification consists of processing input and output signals taken from a system in order to compute an estimate of the parameters defining the mathematical model representing the behavior of the plant under consideration. Thus, the information contained in these signals (input and output) has a strong bearing on the success of the ID algorithm in evaluating the unknown parameters. When the said information content is high, we speak of "good excitation."

In the following sections, we investigate the types of input sequences which maximize the information content of the output of a plant, subject to constraints which make physical sense. The approach taken involves the use of shaping filters which, when driven by white noise, provide their output as the input to the unknown plant. First, we consider fully deterministic and known plants. Although perfect knowledge of the plant obviates the need for ID, we first consider fully known plants

65
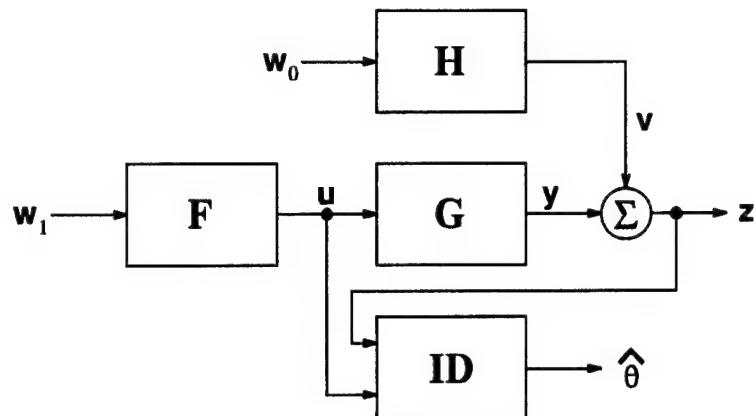
Figure 12. Block Diagram for Open-Loop Identification

in order to gain insight into the problem. Next, we apply these insights to the derivation of optimal shaping filters to be used in the identification of unknown plants. Of course, some prior information about the systems under test is required, such as the standard LTI assumptions, plant order, and *a priori* parameter statistics,.

4.5.1 Information Theory Applied to System Identification. Before we attack the actual shaping filter design, we must interpret some of the concepts discussed in the previous sections within the context of a System Identification thought experiment. Consider the block diagram presented in Figure 12. The plant is designated by **G**, with the ID algorithm accepting inputs to **G** and noise-corrupted outputs from **G**. The measurement noise, **v**, is modeled as the output of **H** driven by $\mathbf{w}_0$. **H** is a shaping filter allowing the modeling of colored measurement noise and $\mathbf{w}_0$ is a unity-strength, normal, white process. Normally distributed noise with unity strength ($\mathbf{w}_1$) drives **F**, the shaping filter which conditions the inputs. $\mathbf{w}_1$ is statistically independent of $\mathbf{w}_0$.

As the diagram shows, the identification uses two signals, namely **u** and **z**. The ID block uses these quantities to generate an estimate of the plant's parameters. In order to facilitate the best (in some sense of the word) estimates, we wish to maximize the information about **G** contained in the output. Although ID requires both input,

u, and output, **z**, the information content in **z** is germane since this is the signal which is modified by the plant under consideration. Furthermore, in this work, **z** is the only quantity subjected to the communications channel; that is, **z** is noise-corrupted.

Thus, we are concerned with maximizing the mutual information about **G** contained in **z**. Recalling Definition 4.2.6 and Equation (85), our goal is to maximize $\mathcal{I}(\mathbf{z}, \mathbf{G})$ where

$$\mathcal{I}(\mathbf{z}, \mathbf{G}) = \mathcal{H}(\mathbf{z}) - \mathcal{H}(\mathbf{z}|\mathbf{G}) \tag{96}$$

Careful inspection of Equation (96) leads to a significant simplification. The conditional entropy term – $\mathcal{H}(\mathbf{z}|\mathbf{G})$ – reduces to a constant, independent of the input. That is, if the plant and its input are known, then the freedom left in **z** is a function only of the measurement noise, **v**. Now, since the $\mathcal{H}(\mathbf{z}|\mathbf{G})$ term is independent of input, maximizing the mutual information via input design is equivalent to maximizing $\mathcal{H}(\mathbf{z})$. Also, in the ID experiment, the actually realized input is known. A complete discussion of this simplification is given by Riggins [75].

4.5.2 Assumptions. Before we can attack the actual shaping filter design, we must clarify the underlying assumptions concerning the plant and measurement noise. First, the plant is assumed stable, linear, and time-invariant. Thus, any steady-state output process resulting from a stationary input process will itself be stationary. Furthermore, while the plant may be non-minimum phase, we cannot allow any zeros of the plant's transfer function to lie directly on the unit circle in the Z-plane. Next, the measurement noise is assumed to be stationary, allowing this noise to be modeled as the steady-state output of a linear, time-invariant, and stable filter driven by white noise.

4.5.3 Shaping Filter Design. With the goal of maximizing the mutual information between the plant and the output established, we proceed to discuss the shaping filter (see Figure 12). System Identification of dynamical systems requires

67

using a history of input and output sequences. Thus, we must maximize the entropy of the output history. In other words, $\mathcal{H}(\mathbf{z})$ alluded to above involves a very long vector containing the output history. Since this history consists of the realizations of a process, it makes sense to consider entropy rate, described in Definition 4.3.3.

Although our aim is to maximize the entropy in the measured output of the plant, we must accomplish this with some restraint; the mathematically formulated entropy maximization problem must make physical sense. Specifically, we shall consider three scenarios. First, we apply a hard constraint on the average output power of the plant. That is, $\mathbf{y}$ must be constrained to some finite variance (which equates to finite average power for a zero mean process). Otherwise, the optimal input would be unbounded, effectively increasing the SNR to infinity. Following similar reasoning, we derive the optimal filter for a limited average input power. Finally, we abandon hard constraints; rather, we apply negative weights to the input and output powers in the performance functional, thereby asking for maximum performance tempered with consideration for the excitation of the plant.

4.5.3.1 Output Power Constraint.

As stated previously, our first goal is to maximize the entropy of the measured output while constraining the output power. So, our problem is stated as a constrained optimization problem:

$$\max_{\mathbf{F}} \overline{\mathcal{H}}(\mathbf{z}) \tag{97}$$

subject to

$$K_y = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n} y_k^2 \tag{98}$$

where we recall that y is the output of the plant, while $\mathbf{z}$ refers to the *measured* output (including measurement noise). The performance functional to maximize and the constraint, given in Equations (97) and (98), are expressed in the time domain. Controls engineers are often comfortable working in the frequency domain. Furthermore, both quantities require an infinitely long sequence of data, which makes the

frequency domain attractive. Hence, we easily transform both into the frequency domain. First, the constraint (98) is expressed in terms of the PSD of the measurement sequence:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n} y_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{S}_{\mathbf{yy}}(\omega) d\omega \tag{99}$$

where we have assumed that $\mathbf{y}$ is a stationary, zero-mean process. Furthermore, Theorem 4.3.2 provides a means to move the performance functional (97) into the frequency domain:

$$\max_{\mathbf{F}} \overline{\mathcal{H}}(\mathbf{z}) = \max_{\mathbf{F}} \left\{ \sqrt{2\pi e} + \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln \left(\mathcal{S}_{\mathbf{zz}}(\omega) \right) d\omega \right\} \tag{100}$$

Now, assuming that $\mathbf{F}$ and $\mathbf{H}$ are both LTI systems driven by Gaussian white noise with unity strength, we can express the performance functional in terms of the magnitudes of the transfer functions:

$$\overline{\mathcal{H}}(\mathbf{z}) = \sqrt{2\pi e} + \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln \left(|\mathbf{F}(\omega)|^2 |\mathbf{G}(\omega)|^2 + |\mathbf{H}(\omega)|^2 \right) d\omega \tag{101}$$

Before attacking the optimization problem, let us examine the concavity of the performance functional.

Lemma 4.5.1 *The functional given in Equation (101) is concave with respect to* $|\mathbf{F}(\omega)|$.

Proof: It is sufficient to show that

$$\ln \left(|\mathbf{F}(\omega)|^2 |\mathbf{G}(\omega)|^2 \right) = \ln \left(|\mathbf{F}(\omega)|^2 \right) + \ln \left(|\mathbf{G}(\omega)|^2 \right)$$

is concave. That is, we must show that

$$\ln \left\{ [a|\mathbf{F}_1(\omega)| + (1-a)|\mathbf{F}_2(\omega)|]^2 \right\} \geq a \ln \left\{ |\mathbf{F}_1(\omega)|^2 \right\} + (1-a) \ln \left\{ |\mathbf{F}_2(\omega)|^2 \right\}$$

$$\forall \, a \in (0,1) \text{ and } \forall \, \mathbf{F}_1, \mathbf{F}_2 \in L_2(-\pi, \pi)$$

Note that $\mathbf{G}$ does not contribute to the concavity. Now, employing the concavity of the natural log function:

$$
\begin{aligned}
\ln\left\{[a|\mathbf{F}_1(\omega)| + (1-a)|\mathbf{F}_2(\omega)|]^2\right\} &= 2\ln\left\{a|\mathbf{F}_1(\omega)| + (1-a)|\mathbf{F}_2(\omega)|\right\} \\
&\geq 2a\ln|\mathbf{F}_1(\omega)| + 2(1-a)\ln|\mathbf{F}_2(\omega)| \\
&= a\ln\left\{|\mathbf{F}_1(\omega)|^2\right\} + (1-a)\ln\left\{|\mathbf{F}_2(\omega)|^2\right\} \\
&\quad\blacksquare
\end{aligned}
$$

We now can derive the optimal shaping filter for maximum information content. Since we are using an equality constraint, the problem of finding an extremum for the performance functional is solved via Lagrange multipliers. To reiterate, our goal is

$$
\max_{\mathbf{F}} \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln\left(|\mathbf{F}(\omega)|^2|\mathbf{G}(\omega)|^2 + |\mathbf{H}(\omega)|^2\right) d\omega \tag{102}
$$

subject to the output power constraint

$$
\begin{aligned}
K_{\text{out}} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{S}_{\mathbf{yy}}(\omega) d\omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} |\mathbf{F}(\omega)|^2|\mathbf{G}(\omega)|^2 d\omega \tag{103}
\end{aligned}
$$

where the performance functional (102) is simplified by ignoring the constant term, and the constraint equation is modified to use the magnitudes of the transfer functions.

First, we form the augmented functional (using the optimization notation adopted from Kirk [38]):

$$
g_a(\omega) = \frac{1}{4\pi} \ln\left\{|\mathbf{F}(\omega)|^2|\mathbf{G}(\omega)|^2 + |\mathbf{H}(\omega)|^2\right\} - \lambda\frac{1}{2\pi}|\mathbf{F}(\omega)|^2|\mathbf{G}(\omega)|^2 \tag{104}
$$

where λ is the Lagrange multiplier.

Next, we set the variation (with respect to $|\mathbf{F}(\omega)|$) equal to zero:

$$
\begin{aligned}
\frac{\partial g_a}{\partial |\mathbf{F}(\omega)|} &= \frac{1}{4\pi} \frac{2|\mathbf{F}(\omega)||\mathbf{G}(\omega)|^2}{|\mathbf{F}(\omega)|^2|\mathbf{G}(\omega)|^2 + |\mathbf{H}(\omega)|^2} - \lambda \frac{1}{\pi} |\mathbf{F}(\omega)||\mathbf{G}(\omega)|^2 = 0 \\
&\Rightarrow |\mathbf{F}(\omega)||\mathbf{G}(\omega)|^2 \left[1 - 2\lambda \left(|\mathbf{F}(\omega)|^2|\mathbf{G}(\omega)|^2 + |\mathbf{H}(\omega)|^2 \right) \right] = 0 \\
&\Rightarrow |\mathbf{F}(\omega)|^2 = \frac{\frac{1}{2\lambda} - |\mathbf{H}(\omega)|^2}{|\mathbf{G}(\omega)|^2}
\end{aligned} \tag{105}
$$

Clearly, $|\mathbf{F}(\omega)| = 0$ is not the solution which makes sense, so we ignore it. Substituting the result of Equation (105) into Equation (103), we have

$$
\begin{aligned}
K_{\text{out}} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{1}{2\lambda} - |\mathbf{H}(\omega)|^2 \right) d\omega \\
&= \frac{1}{2\lambda} - \frac{1}{2\pi} \int_{-\pi}^{\pi} |\mathbf{H}(\omega)|^2 d\omega
\end{aligned} \tag{106}
$$

Solving for the Lagrange multiplier, we have

$$
\begin{aligned}
\frac{1}{2\lambda} &= K_{\text{out}} + \frac{1}{2\pi} \int_{-\pi}^{\pi} |\mathbf{H}(\omega)|^2 d\omega \\
&= K_{\text{out}} + \sigma_v^2
\end{aligned} \tag{107}
$$

where σ_v^2 is the variance of the measurement noise.

The final solution is found by combining Equations (105) and (107):

$$
|\mathbf{F}(\omega)|^2 = \frac{1}{|\mathbf{G}(\omega)|^2} \left[K_{\text{out}} + \sigma_v^2 - |\mathbf{H}(\omega)|^2 \right] \tag{108}
$$

At this point, we can make use of Lemma 4.5.1. By this Lemma, the extremum (if it exists) is indeed a maximum. We address the possible situation in which the solution does not exist in the sequel.

We can make some interesting observations about the optimal shaping filter by careful examination of Equation (108). First, the shape of the filter's frequency response is inversely proportional to the plant's. Thus, input energy is concentrated in frequencies for which the plant exhibits small gain. Furthermore, if the measurement noise is assumed to be colored, the filter tends to attenuate excitation in those

71

frequencies which exhibit high noise. That is, the filter does not 'waste' its energy budget in frequency bands where noise dominates.

Inspection of Equation (108) also indicates a possible problem. We have not restricted the space from which $|\mathbf{F}|$ comes. Therefore, a non-judicious choice of the constraint constant Ko along with the assumed measurement noise model can result in a situation for which the solution does not exist. Specifically, if any frequency exists for which the noise PSD exceeds the sum of the noise variance and the limit on output variance, then the solution calls for an imaginary magnitude. An example situation in which this problem could exist is a noise model which includes a large, narrow spike in the frequency domain, such as would occur if the designer expects a relatively strong sinusoidal component in the measurement noise. Obviously, this type of situation is intractable given the method of solution used in this derivation. If the designer is adamant about these choices of the noise model and the constraint, more sophisticated constrained optimization methods must be employed. The available methods of solution involve complicated and problem-specific choices. This type of situation is not further addressed by this research, other than to alert the reader that the possibility exists.

Of course, there may be *ad hoc* methods for addressing the problem in which the above optimal solution does not exist. For example, if the optimal solution includes a small frequency band in which the input PSD is negative, we could approximate the curve by setting the PSD to zero (or near zero) for these frequencies. This type of suboptimal solution would result in an input which violates the constraint, but a simple gain adjustment could remedy this particular problem.

In summary, we have derived an optimal solution by which information content in the measurements is maximized while conforming to a constraint on the plant's output power.

While the output power is probably of primary interest (especially in an online application), we may also wish to restrict input power. Input power constraints

would be most applicable when the plant includes some inherent limitation on actuator power. The next section addresses the optimization under input constraints.

4.5.3.2 Input Power Constraints. We now calculate the optimal shaping filter which maximizes output entropy while constraining the input power to some fixed value given by K_{in}. The performance functional is unchanged from Equation (101) but the constraint is modified to account for the input power to the plant. We state the optimization as:

$$\max_{|\mathbf{F}|} \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln\left(|\mathbf{F}(\omega)|^2 |\mathbf{G}(\omega)|^2 + |\mathbf{H}(\omega)|^2\right) d\omega \tag{109}$$

subject to

$$
\begin{aligned}
K_{\text{in}} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{S}_{\mathbf{uu}}(\omega) d\omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} |\mathbf{F}(\omega)|^2 d\omega
\end{aligned}
\tag{110}
$$

The method of solution is similar to that used in the previous section. That is, we incorporate a Lagrange multiplier and set the total variation to zero, finding a necessary condition for an extremum. Again, by Lemma 4.5.1, if the extremum exists, it will be a maximum. Moving directly to the solution, we have

$$|\mathbf{F}(\omega)|^2 = K_{\text{in}} + \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|\mathbf{H}(\omega)|^2}{|\mathbf{G}(\omega)|^2} d\omega - \frac{|\mathbf{H}(\omega)|^2}{|\mathbf{G}(\omega)|^2} \tag{111}$$

This solution also exhibits some interesting properties. First, if the measurement noise is ignored ($\mathbf{H}(\omega) = 0$), the shaping filter collapses to a constant. In other words, the maximum entropy at the output of the plant is achieved by white noise input (under input power constraint). This situation also occurs if we allow very large excitation (i.e. K_{in} very large), allowing the filter to employ a large SNR. However, if we account for measurement noise, the flavor of the optimal input changes markedly. For example, a simple white measurement noise model yields an input PSD which is somewhat attenuated in frequency intervals for which the plant has low gain and is amplified for those frequencies for which the plant exhibits high gain.

As is the case for the filter designed under output power constraint, this filter tends to allocate its energy in frequencies for which noise strength is low. However, in contrast to the previous solution, input power constraints allow the input to place energy in plant-high-gain frequencies.

Unfortunately, this solution may not exist in an implementable form. If our models of the plant and the measurement noise interact such that $|\mathbf{F}(\omega)|^2 < 0$ for some frequency, viz., an infeasible solution results, then the choice of simple Lagrange multipliers must be abandoned for a more sophisticated constrained optimization technique which allows the imposition of hard (inequality) constraints, such as $|\mathbf{F}(\omega)|^2 \geq 0$. As is the case with the output power constrained case, we would be forced either to apply problem-specific approaches to the optimization involving duality theorems as covered by Luenberger [46], or apply suboptimal approximations to the optimal solution.

4.5.3.3 Input and Output Power Penalties. Our previous two approaches allow us to design an optimal shaping filter satisfying constraints on the input power with no regards to the output power and vice versa. If we are concerned with both quantities, we could employ simultaneous inequality constraints on both input and output power. However, inequality constraints will require solution methodologies which are, again, problem specific and make the posed optimization problem difficult. Hence, we are motivated to adopt a different approach in this case. Namely, in the spirit of Linear Quadratic Optimization, we form a new performance/cost functional which employs positive weighting to the information rate of the output while penalizing input and output power. The following functional is considered:

$$
\begin{aligned}
J(|\mathbf{F}|) = & \int_{-\pi}^{\pi} \left[\frac{1}{2} \ln \left(|\mathbf{F}(\omega)|^2 |\mathbf{G}(\omega)|^2 + |\mathbf{H}(\omega)|^2 \right) \right] d\omega \\
& - \int_{-\pi}^{\pi} W_y |\mathbf{F}(\omega)|^2 |\mathbf{G}(\omega)|^2 d\omega - \int_{-\pi}^{\pi} W_u |\mathbf{F}(\omega)|^2 d\omega
\end{aligned}
\tag{112}
$$

where W_u and $W_y \geq 0$ are designer-chosen weights. Clearly, this functional reflects the aforementioned goals. The first term of the integrand is identified as the entropy contribution to the functional, while the second and third terms apply negative weight to the output and input powers, respectively. Furthermore, Lemma 4.5.2 shows that this functional is concave with respect to $|\mathbf{F}|$.

Lemma 4.5.2 *The functional given in Equation (112) is concave with respect to $|\mathbf{F}(\omega)|$, providing the weights identified by W_u and W_y are non-negative.*

Proof: Lemma 4.5.1 yields concavity of the first term. The second and third terms are negatively weighted quadratic terms in $|\mathbf{F}|$; thus they are concave. Now, a sum of concave functionals is concave, so the entire functional is concave. ∎

Thus, an extremum (if one exists) represents a maximum of the functional.

By construction, this optimization is unconstrained, obviating the need for Lagrange multipliers. Thus, we set the total variation (with respect to $|\mathbf{F}|$) equal to zero:

$$\frac{|\mathbf{F}(\omega)||\mathbf{G}(\omega)|^2}{|\mathbf{F}(\omega)|^2|\mathbf{G}(\omega)|^2 + |\mathbf{H}(\omega)|^2} - 2W_y|\mathbf{F}(\omega)||\mathbf{G}(\omega)|^2 - 2W_u|\mathbf{F}(\omega)| = 0 \qquad (113)$$

which yields the expression for the optimal shaping filter's magnitude:

$$|\mathbf{F}(\omega)|^2 = \frac{1}{2W_y|\mathbf{G}(\omega)|^2 + 2W_u} - \frac{|\mathbf{H}(\omega)|^2}{|\mathbf{G}(\omega)|^2} \qquad (114)$$

Once again, an implementable $|\mathbf{F}(\omega)|$ (≥ 0) may not ensue for certain choices of the weights and the measurement noise model. However, assuming a feasible solution exists, the interaction between the weights W_u, W_y, and the noise model is complicated and interesting.

First, consider the situation in which very small weighting (penalty) is applied to the input and output power. In this case, the first term dominates, driving the shaping filter magnitude very high. Furthermore, the filter's magnitude indicates a tendency to attenuate the input energy within frequencies for which the plant has
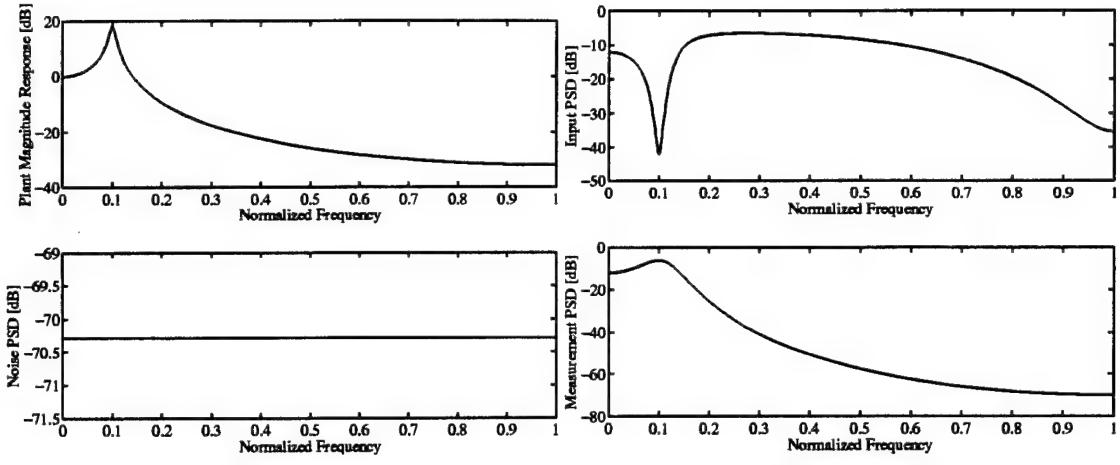
(a) Plant and Noise Model   (b) Input and Output PSD

Figure 13.   Input and Output PSD *resulting from the unconstrained optimal shaping filter designed to a low-pass plant with lightly damped poles and white measurement noise.*

high gain. In those frequencies for which $|\mathbf{G}(\omega)|$ is low, the optimal input tends to look like high-variance white noise.

Notice also, that if either weight is set to zero, the resulting solution adopts a form similar to those given in Equations (108) or (111). This is not unexpected since the method of Lagrange multipliers reduces a constrained optimization problem into an unconstrained form.

Finally, consider the case in which both terms are active. For example, Figure 13(a) illustrates the magnitude response for a plant with a set of lightly damped poles. The output is corrupted by white noise. With the weights set equal to each other, Figure 13(b) shows the optimal input PSD and the resulting output PSD. Notice how the optimal input is notched near the plant's resonant peak near the normalized frequency 0.3. We also note that the optimal input PSD drops off in high frequencies. Thus, the optimal filter behaves as the inverse of the plant where the plant has high gain. However, the filter's frequency response is shaped similarly to the plant's in the higher frequencies where the plant gain falls off. Our next

76

(a) Plant and Noise Model   (b) Input and Output PSD
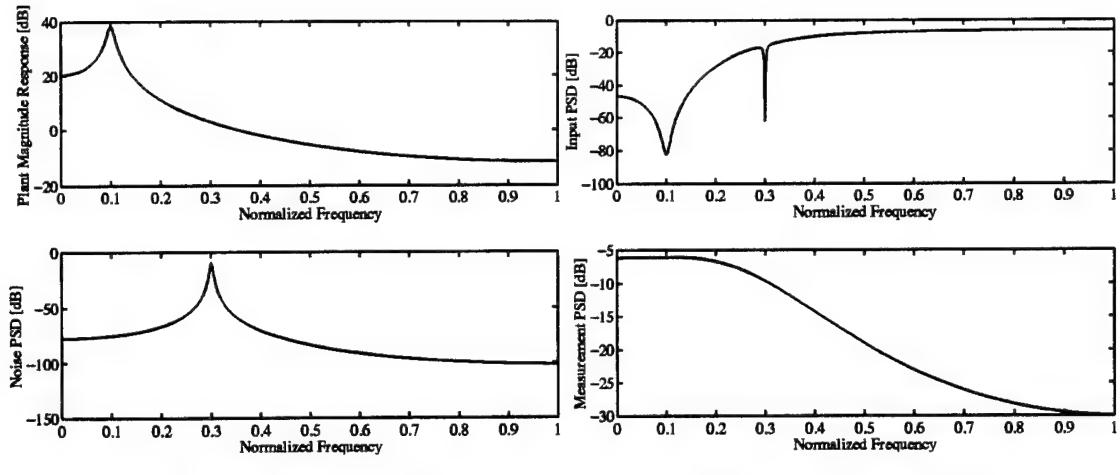
Figure 14. Input and Output PSD *resulting from the unconstrained optimal shaping filter designed to a low-pass plant with lightly damped poles and highly resonant measurement noise.*

experiment, illustrated in Figures 14(a) and 14(b), uses the same plant, but with a noise model incorporating a sharp, narrow peak. This type of noise model includes a sinusoidal component in the measurement noise. Figure 14(b) shows the same inverse-plant behavior, but with a notch near the resonant peak of the measurement model. Thus, the optimal input excludes excitation in the frequencies for which noise dominates.

4.5.4 Output Power Constraint – Stochastic Formulation.   The previous sections provide the derivations for the optimal input PSD given perfect knowledge of the plant. Although these preliminary formulations yielded useful insights into the problem, the practical engineer requires an optimal input for ID when faced with uncertain plants. At this point it is worth mentioning that our results should be applied in an iterative way; thus, the current plant estimate is used to design an optimal input to be applied to the plant, following which the ID algorithm is applied to the input/output pair and an improved plant estimate is obtained. Nevertheless, we shall see in the sequel that we can directly include plant uncertainty and arrive

77

at a solution for the optimal shaping filter by replacing every instance of the plant's square magnitude with the expected value of the same. We will concentrate on the output-power constraint formulation for two reasons. First, we are most concerned with limiting the plant's response to the probing input in an on-line application. Second, the development for the stochastic formulation of the input-power constraint and the unconstrained cases closely parallels the development offered here.

Recalling Equations (102) and (103), our goal was

$$\max_{\mathbf{F}} \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln\left(|\mathbf{F}(\omega)|^2 |\mathbf{G}(\omega)|^2 + |\mathbf{H}(\omega)|^2\right) d\omega$$

subject to

$$K_{\text{out}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\mathbf{F}(\omega)|^2 |\mathbf{G}(\omega)|^2 d\omega$$

Now that we are considering uncertain plants, we restate the problem with expected values. That is, our optimization problem now is

$$\max_{\mathbf{F}} \mathcal{E}_{\boldsymbol{\theta}}\left\{\frac{1}{4\pi} \int_{-\pi}^{\pi} \ln\left(|\mathbf{F}(\omega)|^2 |\mathbf{G}(\omega)|^2 + |\mathbf{H}(\omega)|^2\right) d\omega\right\} \tag{115}$$

subject to the expected output power constraint

$$K_{\text{out}} = \mathcal{E}_{\boldsymbol{\theta}}\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} |\mathbf{F}(\omega)|^2 |\mathbf{G}(\omega)|^2 d\omega\right\} \tag{116}$$

In other words, we wish to maximize the expected entropy of the output subject to a constraint on the expected average output power.

First consider the constraint, Equation (116). Expectation and integration are both linear operators, so we pass the expectation operator into the integral:

$$K_{\text{out}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\mathbf{F}(\omega)|^2 \mathcal{E}_{\boldsymbol{\theta}}\left\{|\mathbf{G}(\omega)|^2\right\} d\omega \tag{117}$$

where we have made the tacit assumption that $\mathcal{E}_{\boldsymbol{\theta}}\{|\mathbf{G}(\omega)|^2\}$ exists $\forall \omega \in [-\pi, \pi]$; a discussion of the existence problem is given in Appendix B.

Similarly, we can pass the expectation operator into the integral in the performance functional, Equation (115). Furthermore, the log function is concave; i.e.

$$\log\left[\mathcal{E}_{\boldsymbol{\theta}}\left\{|\mathbf{F}(\omega)|^2\mathbf{G}(\omega)|^2\right\}\right] \geq \mathcal{E}_{\boldsymbol{\theta}}\left\{\log\left[|\mathbf{F}(\omega)|^2\mathbf{G}(\omega)|^2\right]\right\} \tag{118}$$

With Equation (118) in mind, we pose the new optimization problem

$$\max_{\mathbf{F}} \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln\left[|\mathbf{F}(\omega)|^2\mathcal{E}_{\boldsymbol{\theta}}\left\{|\mathbf{G}(\omega)|^2\right\} + |\mathbf{H}(\omega)|^2\right] d\omega \tag{119}$$

subject to the constraint given in Equation (117), and assuming the existence of $\mathcal{E}_{\boldsymbol{\theta}}\{|\mathbf{G}(\omega)|^2\}$. Thus, we are maximizing an upper bound of the original performance functional given by Equation (115).

The optimal shaping filter's magnitude for this stochastic formulation is derived in a manner identical to the development of Equation (108). We find that the filter which satisfies the stochastic formulation is described explicitly by

$$|\mathbf{F}(\omega)|^2 = \frac{1}{\mathcal{E}_{\boldsymbol{\theta}}\{|\mathbf{G}(\omega)|^2\}}\left[K_{\text{out}} + \sigma_v^2 - |\mathbf{H}(\omega)|^2\right] \tag{120}$$

Thus, the optimal input's PSD is completely described by the product of a term wholly dependent on the available plant prior information and a second term derived only from the known measurement noise model.

4.5.5 Implementation Issues. Equations (108), (111), and (114) provide us with elegant expressions for the optimal shaping filter's magnitude squared. Since we began the derivation by assuming the filter is driven by unity-strength Gaussian white noise, these expressions are interpreted as the filter output's PSD. Aside from the feasibility problems already mentioned, we must be concerned with implementing the shaping filters as difference equations. A particular difficulty is caused by the negative sign which occurs in each filter's spectrum. Although not insurmountable, this complication requires that we do some careful manipulation. In fact, we may be faced with the situation in which the input PSD cannot be realized with a finite dimensional filter. Here, we would be forced to use a suboptimal approxi-

mation, such as a numerical fitting. This approximation could be accomplished via implementation of the modified Yule-Walker equations [21].

In particular, the optimal filter based on output power constraints should be considered for implementation. We are motivated to use this formulation for two reasons. First, the situation in which we wish to limit output power (and thus, response of the plant to excitation) is most appropriate for many applications. For example, if the excitation used for identification significantly affects flying qualities of an aircraft, then the signal is producing an output with too much output power. The form of the solution provides the second motivation for using the output power constraint formulation. That is, the solution is most amenable to on-line adaptivity. Recalling Equation (108):

$$|\mathbf{F}(\omega)|^2 = \frac{1}{|\mathbf{G}(\omega)|^2} \left[K_{\mathrm{out}} + \sigma_v^2 - |\mathbf{H}(\omega)|^2 \right]$$

we see that the required PSD is implementable as the output of two cascaded filters driven by white noise. These two filters require magnitude responses equal to the plant's inverse and a function of the difference between a constant and the noise PSD. The first filter in line is easily given by a simple inversion of the plant's transfer function with enough delays added to make the inverted transfer function proper. The next filter can be implemented with a numerical solution. The beauty of this approach is that the numerically derived portion of the solution is a function only of the noise model; thus, it does not change as the plant estimate changes. So if we wish to adapt the input signal to reflect the latest estimate of the plant, we need only to invert the estimated transfer function of the plant.

One particularly attractive implementation of optimally-colored input for System ID is linked to Multiple Model Adaptive Estimation (MMAE) [53]. An MMAE estimator can be configured to generate a vector of probabilities, each assigned to one of a set of possible models. In other words, we take the parameter vector to be a discretely-distributed random vector. Each possible value for the parameter vector

80

yields a different model, so the MMAE selects the 'correct' model from a finite set. Now, each model in the set has associated with it an optimal input and a probability that the model matches the plant. Thus, we can use these probabilities to generate a weighted blend of the optimal inputs. Furthermore, as we note in Appendix B, the existence problem is obviated by the use of a discrete set of possible plants.

The MMAE-like concept is illustrated in Figure 15. The primary components illustrated here are:

- $\mathbf{G}$ is the plant under test.

- $\mathbf{H}$ is the measurement noise coloring filter (typically a simple gain yielding white noise).

- $\mathbf{F}_\mathrm{H}$ is part of the input-shaping filter. This filter is derived from the measurement noise model.

- The ID block is MMAE with outputs $p_1, \ldots, p_M$ (i.e. the probabilities associated with M different stipulated models.)

- $\mathbf{F}_1, \ldots, \mathbf{F}_M$ are shaping filters associated with each of the hypothesized M models.

- $q_1, \ldots, q_M$ are weights to be applied to each shaping filter's output.

- The algorithm block conditions the probabilities and produces values for the weights, $q_1, \ldots, q_M$.

- $\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_M$ are independent random sequences, each of which is unity strength, white, and Gaussian.

First, we derive $\mathbf{F}_\mathrm{H}$ from the second term of Equation (120):

$$\mathbf{F}_\mathrm{H}(\omega) = \left[K_\mathrm{out} + \sigma_v^2 - |\mathbf{H}(\omega)|^2 \right] \tag{121}$$

where σ_v^2 is the variance of the measurement noise, computed by

$$\sigma_v^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\mathbf{H}(\omega)|^2 d\omega$$

81
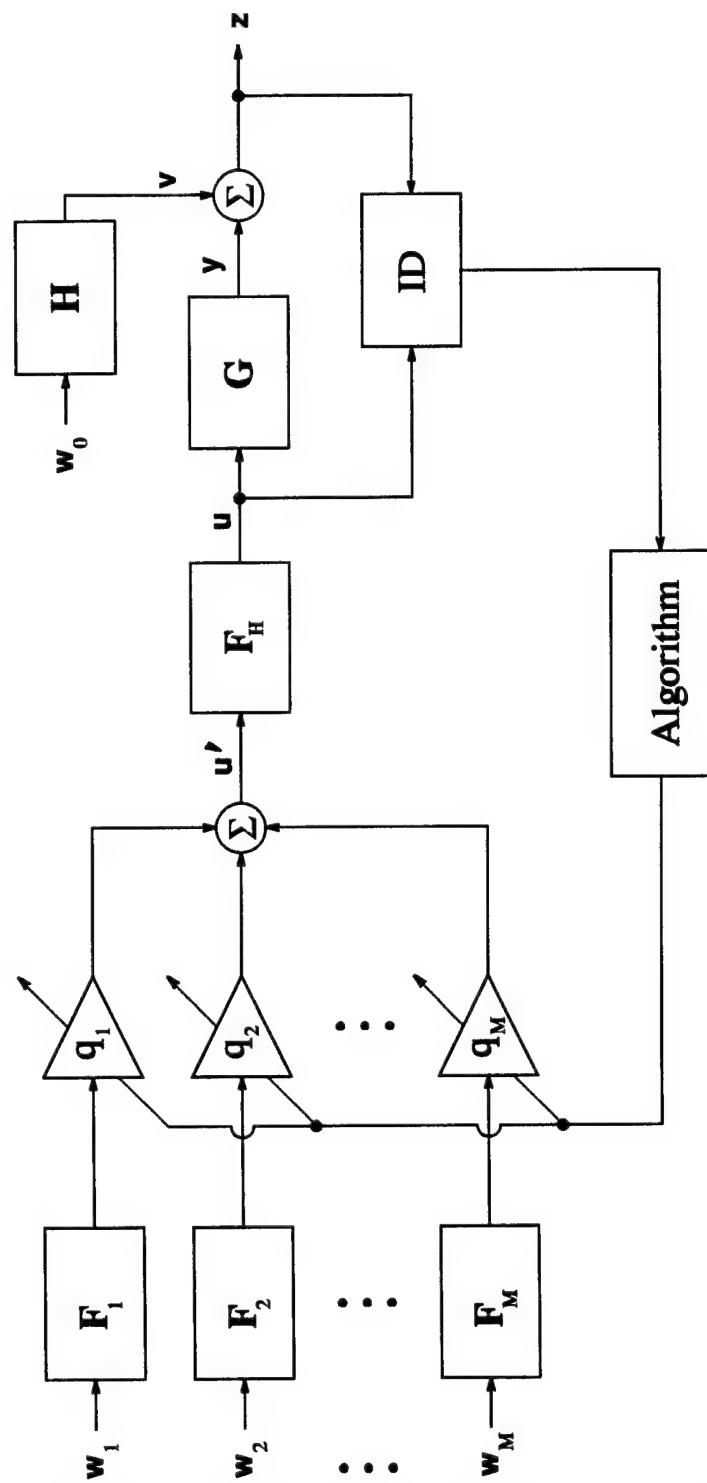
Figure 15. Implementation of the Optimally Colored Input in a MMAE Setting

Again, we are assuming that $\mathbf{w}_0$ is unity strength white Gaussian noise. As we noted previously, if we assume that the measurement noise is white, then $\mathbf{F}_H$ collapses to a constant. However, if the measurement noise is colored, then the magnitude of $\mathbf{F}_H$ takes on a more complex shape in the frequency domain. If necessary, we can design a filter with the proper shape (or arbitrarily close to the proper shape) via the use of many filter design packages. For example, the Matlab® Signal Processing Toolbox offers the *yulewalk* command which incorporates the modified Yule-Walker equations.

Now we need to generate the signal labeled $\mathbf{u}'$ such that its PSD is described by $\mathcal{E}_{\boldsymbol{\theta}}\{|\mathbf{G}(\omega)|^2\}^{-1}$. The optimal solution here is to form a filter with a magnitude characteristic described by

$$|\mathbf{F}_{\text{optimal}}(\omega)|^2 = \left[\mathcal{E}_{\boldsymbol{\theta}}\{|\mathbf{G}(\omega)|^2\}\right]^{-1} = \left[\sum_{i=1}^{M} p_i |\mathbf{G}_i(\omega)|^2\right]^{-1} \qquad (122)$$

The expectation is taken over the finite number of models inherent to the MMAE approach, where p_i is the probability that the ith model is correct.

The expression given by Equation (122) presents some problems. First, we will not, in general, be able to realize a single filter with the specified magnitude characteristic. Next, if the filter is realizable, we still need to redesign each time the MMAE generates a new set of probabilities; this would require an undesirable level of computation in an on-line setting.

Fortunately, an MMAE often selects one model by assigning a high probability to that model and very low probabilities to each of the other models [53]. If we assume that this will be the normal operation for the estimator, then the algorithm block becomes a simple selector, setting $q_i = 1$ where the ith model is the most correct and $q_j = 0$ for all other models. We also design each of the elemental shaping filters to have a magnitude characteristic equal to the inverse of the corresponding model. Thus, the input is switched to be optimal for the currently selected plant model. The switching between one input and another will introduce high frequency

83

transients into the input stream. Hence, we might also wish to include a bank of filters in the algorithm block to smooth the transitions between inputs.

Another approach for determining the weights q_j is to allow the 'algorithm' block to simply pass the square root of the probabilities to the corresponding weight. That is, $q_j = \sqrt{p_j}$ for $j = 1, \ldots, M$. Given this approach, the PSD of $\mathbf{u}'$ is

$$\mathcal{S}_{\mathbf{u}'\mathbf{u}'}(\omega) = \sum_{i=1}^{M} \frac{p_i}{|\mathbf{G}_i(\omega)|^2} \tag{123}$$

This formulation is certainly suboptimal, but as any one probability approaches unity, the input PSD approaches the optimal.

4.6 Conclusions

In this chapter, an area of information theory which is usually applied to communications is explored in the context of optimal experiment design for System ID. Extending an analogy between communications and identification experiments, we used entropy to quantify the level of information contained in output measurements. Although the term 'information' can take on dichotomous meanings, we saw that we can interpret the information content as a quantity to be maximized within the context of a System Identification experiment. Therefore, we adopted output entropy as an ID enhancement performance functional. Furthermore, entropy can be calculated using the power spectral density of a stationary signal, which lead us to maximizing the information in a plant's output via the use of optimal input shaping filters driven by unity-strength white noise.

In order to limit the excitation of the plant, we formulated three different constrained optimization problems for the derivation of the optimal shaping filter. In all three formulations, our goal was to maximize output entropy while tempering the excitation with constraints or penalties. The formulations make physical sense and the ensuing optimal-input-generating filters are obtained. The first formulation involves a constraint on the plant's average output power. Second, we optimize

on information content, constrained by the plant's average input power. Our final formulation employs penalties on both input and output power with no constraints.

We first constrained output power and found that the optimal filter's magnitude is proportional to the inverse of the plant multiplied by a term defined by the measurement noise model. If this model is white noise, then the noise-driven term in the shaping filter's magnitude becomes a constant. We saw that a colored measurement noise model yields a term in the shaping filter's magnitude which exhibits some attenuation of the input in frequencies wherein the noise is high. That is, the optimal filter concentrates more input energy in those frequencies of low noise, avoiding the 'waste' of energy within frequencies where noise might dominate.

Next, we applied a constraint to the average power applied to the input of the plant while maximizing the entropy of the output. We observed that this optimal filter becomes a constant (applying white noise to the input of the plant) when the measurement noise is very small. However, assuming a significant level of measurement noise yields an optimal shaping filter with a more interesting shape. As we saw in the output power constraint case, the optimal filter tends to attenuate the input in those frequencies in which the measurement noise is high and amplify input in frequencies of lower noise. In contrast to the filter designed under output power constraint, this filter's shape tracks the plant's magnitude, amplifying the input in frequencies for which the plant has higher gain.

As a third alternative, we explored an unconstrained formulation of the optimization problem whereby the performance functional includes positive weight on the output entropy while penalizing the plant's input and output power. This approach yields a shaping filter with a complex and interesting shape. We saw that setting the penalties on the input and output power very small results in a filter which approaches the inverse of the plant in magnitude with a large gain, thus forcing the plant's output to look like very high strength white noise. As we expected, if either penalty is set to zero, the optimal shaping filter takes on a magnitude characteristic

similar to one of the constrained formulations. Most interesting, if both penalties are significant, the optimal shaping filter looks like a blend of the two filters derived under constraints. We saw that this filter tends to notch out frequencies for which the plant exhibits very high gain, but tends to roll off in magnitude for frequencies for which the plant's magnitude drops. Thus, the unconstrained filter is shaped as the inverse of the plant near plant-resonance frequencies and takes a shape similar to the plant's magnitude response where the plant's magnitude rolls off.

While each of the three previously discussed optimal filters is based on a deterministic plant model, our fourth formulation includes plant uncertainty. We included uncertainty in the plant's parameters by optimizing on the expected value of the output entropy bound while constraining the expected average plant output power. The optimal shaping filter which results has a form similar to the deterministic case, but with the plant's squared magnitude replaced by the expected value of the squared magnitude.

Finally, we looked at implementation issues. The optimal inputs derived here are described by the input's optimal power spectral density which is not always achievable via the use of realizable filters driven by white noise. We can, however, approximate the required PSD with filters designed to fit the required spectrum through the use of numerical filter-design algorithms. Furthermore, we saw that the inputs derived for uncertain plants can be realized sub-optimally by a set of filters running in parallel and cascaded through another filter. The latter filter is described completely by the measurement-noise model while the former requires knowledge of the current estimate of the plant. This allows us to design the noise-driven filter a priori with no knowledge about the plant. The outputs of the parallel filters are blended through a series of adaptive gain elements, each gain determined by a set of probabilities taken from a Multiple Model Adaptive Estimator. This approach will be investigated fully in future research.

With the theory defining optimal inputs for identification in place, we move to experimental results in Chapter V. As practical engineers, we wish to see the advantage in using the optimally colored inputs described in this chapter. That is, although we have designed entropy (information) maximizing inputs, we wish to observe how this maximum information is manifested in the reduction of the estimated parameter errors when an efficient System ID algorithm is employed. Furthermore, we will see experiments which corroborate the theory presented here.

V. Experiments

In Chapter II the theory relating the order of the input to the identifiability of the parameters defining the math model of a plant was developed. Chapter III addressed the development of efficient identification algorithms in the presence of measurement and process noise. An algorithm for accomplishing Minimum Variance/Iterated Weighted Least Squares identification which dramatically outperforms standard Least Squares estimation was developed and tested. In Chapter IV the calculation of the input frequency spectrum which optimizes the information content of the input/output maximizing the potential for successful ID was addressed. In this chapter, experiments which are designed specifically to provide evidence to investigate the utility of some of the theory presented in Chapter IV are performed. We will collect this evidence by examining the effects different inputs have on ID performance, when the ID algorithm referred to above is used.

This chapter is organized into four main sections. Section 5.1 presents a short discussion on various metrics which are used to quantify successful parameter identification. In Section 5.2 several experiments through which we draw some conclusions about frequency-domain concentration for the excitation are performed. Section 5.2.1 documents an experiment through which we apply inputs consisting entirely of sinusoids of varying frequencies and observe some widely used ID metrics. Sections 5.2.2 and 5.2.3 present the results of experiments which investigate the use of various shaping filters to color the input. Next, Section 5.3 documents the results of some parameter estimation experiments. Finally, Section 5.4 continues a discussion of the experimental results.

5.1 Identification Metrics

We wish to explore the relationship between different types of inputs and the identifiability of the parameters defining a dynamical system. Therefore, we need

a good metric which quantifies the potential for identifiability. The available literature offers several metrics which accomplish this goal by reflecting the parameter estimation error covariance into a scalar; for example see [28, 32, 44, 55, 96]. Under some non-restrictive assumptions on the plant, and assuming an unbiased estimator is available – which, unfortunately, is seldom the case except when ARX models are used – the Fisher Information Matrix $\mathbf{M}$ provides us with the inverse of the parameter estimate error covariance matrix. Thus, the following metrics all involve the use of $\mathbf{M}$ [32]:

- A-optimality: $J_A(\mathbf{M}^{-1}) = \text{trace}(\mathbf{A}\mathbf{M}^{-1}), \quad \mathbf{A} > 0$

- D-optimality: $J_D(\mathbf{M}^{-1}) = \det(\mathbf{M}^{-1})$

- E-optimality: $J_E(\mathbf{M}^{-1}) = \lambda_{\max}(\mathbf{M}^{-1})$ (i.e. the maximum eigenvalue)

- C-optimality: $J_C(\mathbf{M}^{-1}) = \text{cond}(\mathbf{M}^{-1})$

Each of these criteria is minimized for optimality.

J_A is attractive since it is easy to compute and we easily can see the link between A-optimality and the minimization of the estimation error variance. Furthermore, the choice of $\mathbf{A}$ provides a transparent way to place more or less importance on particular elements in the unknown parameter vector. C-optimality is applied to increase the convergence rate of the gradient algorithm. Furthermore, recalling Equation (49):

$$\hat{\boldsymbol{\theta}}_{MV} = \left[\mathbf{H}_{k-1}^{(q)}{}' \mathbf{R}^{-1} \mathbf{H}_{k-1}^{(q)} \right]^{-1} \mathbf{H}_{k-1}^{(q)}{}' \mathbf{R}^{-1} \mathbf{z}_k^{(q)}$$

we see that we require a matrix inversion for Weighted Least Squares parameter estimation. We wish for this matrix to be as far from singular as possible, thus minimizing the sensitivity to errors (noise) in the solution vector. As Maybeck [53] points out, this matrix product is recognized as one form of the Fisher Information Matrix. Now, the condition number of a matrix serves to quantify the sensitivity to noise in the solution vector $\mathbf{z}_k^{(q)}$ (i.e. the 'singularity' of the matrix), so we are motivated to minimize the condition of the Information Matrix.

We might also think of the determinant (used in D-optimality) as measure of the 'singularity' of the matrix. After all, a singular matrix's determinant is zero. Thus, we might be tempted to maximize the determinant of the Information Matrix. However, a large determinant does not imply a small condition number. Thus, we maintain that maximizing the determinant of the Information Matrix does not necessarily ensure a well-conditioned ID problem.

As Herrera-Bendezú [32] points out, minimization of the J_E functional leads to minimizing the maximum radius of the ID error ellipsoid. This makes E-optimality appealing, but it has not received much attention in the literature because there is not an analytic expression for the eigenvalues of a general matrix. Furthermore, we can monitor the anticipated error by observing the trace of the error covariance matrix $(\mathbf{M}^{-1})$.

We are concerned with minimizing the estimation error committed by the ID algorithm in the presence of measurement noise. To this end, we will focus on C-optimality and A-optimality (with $\mathbf{A}$ fixed as the identity matrix, applying equal weight to all parameter variances). We choose to monitor the condition number since it quantifies the sensitivity of the ID algorithm to noise. Furthermore, we will observe the trace of the error covariance matrix as a direct reflection of the spread of the errors in the parameter estimates.

5.2 Input Frequency Concentration

The experiments documented here are designed to address the effects of frequency content of the input. First, we consider inputs comprised of sinusoids restricted within different frequency bands. Next, we use the outputs of several bandpass filters driven by white noise as the input. Finally, we replace the bandpass filters with bandstop filters. In each case, we run the experiment for several locations of the bands, allowing us to compute the ID metrics as a function of the location of the band. A more detailed description of the experiments is provided in the sequel.

Truth Model Response
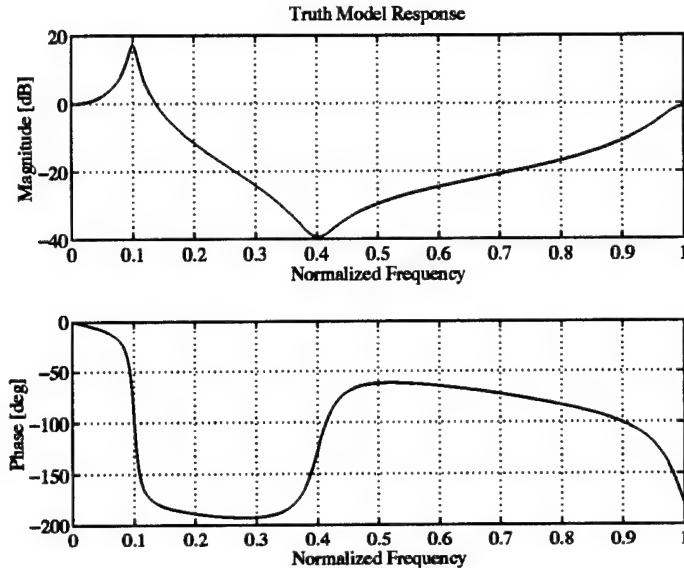
Figure 16. Bode Plots: *This system has a pair of lightly-damped poles and a pair of zeros near the unit circle, producing a resonant peak and an anti-resonant dip.*

The plant which serves as a truth model for these experiments is not rooted in reality. Rather, the plant is designed to focus our attention on portions of the frequency domain in which the plant exhibits high gain plus those frequencies of low magnitude. That is, the plant's magnitude characteristic includes a sharp resonant peak and a sharp anti-resonant dip, as illustrated by the Bode plots shown in Figure 16. The truth-model transfer function is

$$\mathbf{G}(z) = \frac{b_1 z^2 + b_2 z + b_3}{z^3 + a_1 z^2 + a_2 z + a_3} = \frac{0.14323 z^2 - 0.081442z + 0.12123}{z^3 - 0.96407 z^2 - 0.71726z + 0.86436}$$

which has three poles: $-0.9, 0.98\angle\pm0.1\pi$ and two zeros: $0.92\angle\pm0.4\pi$. In particular, notice that the plant's magnitude response does not exhibit high-frequency rolloff. This is not a realistic situation. However, a more realistic plant model is used in experiments documented in following sections of this chapter.

5.2.1 Sinusoidal Inputs. Chapter II presented a transparent way to relate the order of excitation to the model order. We saw that identification of a plant with n unknowns requires an input of at least order n. For example, a second-order

91 91

91

plant (with four unknown parameters) must be excited by an input consisting of linear combinations of at least two sinusoids, i.e., of order four, for identification. Although we know the minimum number of sinusoids required for identification, we have not addressed the best placement (in frequency) of these sinusoids.

In order to investigate the effects of the placement of sinusoidal inputs on ID performance, we conduct the following experiment. A truth-model is excited by an input consisting of a sum of sinusoidal signals, with the magnitudes of the sinusoids chosen to achieve an average output power equal to one. Hence, we are considering the case of constrained output power. (Unity output power is chosen arbitrarily.) The group of sinusoids is grouped tightly around a center frequency. We then build a regressor matrix from the inputs and steady-state outputs. This regressor is then used in calculating the ID-performance metrics. Namely, we consider the condition number of the Information Matrix and the trace of the estimation error covariance matrix.

We determine the best placement of the input in the frequency domain by sweeping the center frequency through the entire allowable range of frequencies. That is, the entire band of frequencies making up the input must lie in (0,1) in Normalized Frequency Units (NFU), where all frequencies are normalized such that unity represents the Nyquist or folding frequency. We calculate and store the metrics corresponding to each center frequency and plot the results.

Figure 17 should help to illustrate the methodology used in this experiment. Each hatched region in this figure represents a band in which the input frequencies lie; the input is taken as a sum of sinusoids having these frequencies. This input is then applied to the plant and the resulting input/output pair is used to construct a regressor. We then calculate and store the desired metrics from the regressor. Having completed this phase of the experiment, we move to the next band and re-accomplish the procedure. Note that the bands illustrated in Figure 17 are designed

92
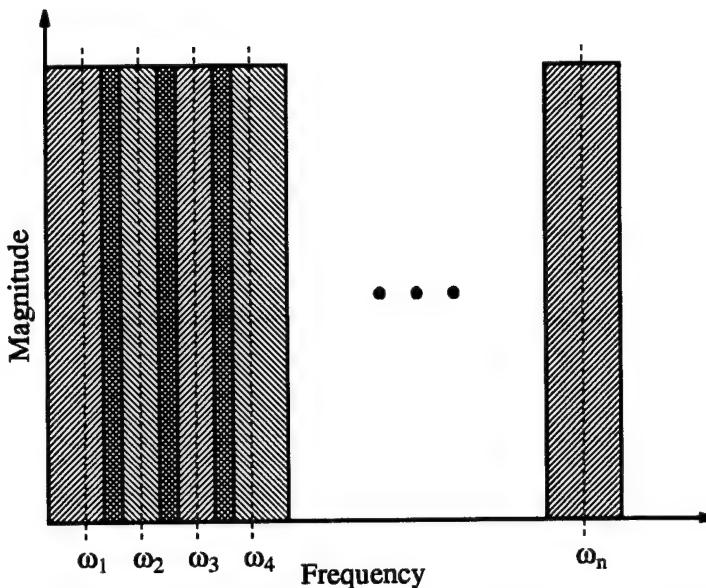
Figure 17. Excitation Bands: *The hatched areas illustrate the bands of frequency in which excitation is concentrated (or attenuated). For the sinusoidal-input experiments, each run of each experiment uses sinusoids with frequencies contained in each of the bands. The bandpass experiments use filters whose ideal magnitude characteristics are illustrated by each of the n bands. In the case of notch (or bandstop) filter experiments, the hatched areas represent the ideal stop-band magnitude characteristic.*

to overlap, allowing the center band-center frequencies (denoted ω_1, ω_2, ...) to be arbitrarily close.

For this experiment, we set the width of each band to 0.05 NFU. We wish the band to be narrow in order to isolate a small region of frequencies, but not so narrow as to induce numerical difficulties in distinguishing frequencies within the band. Hence, our choice of bandwidth is arbitrary, but not without consideration. For each center frequency, we apply an input consisting of a sum of twenty sinusoids equally spaced within the band identified by its center. As we saw in Chapter II, this system with six unknown parameters requires an excitation consisting of at least three sinusoids. This, however, is a minimum. We wish to use a large number of frequencies to mitigate the so-called 'picket-fence effect' [83] by which important information is missed between the frequencies sampled by the sinusoidal input components. Once

again, possible numerical problems motivate us to keep the number of frequencies relatively small. Therefore, we settle on twenty sinusoids as a compromise.

The results of this experiment are shown in Figures 18 and 19. Figure 18(a) illustrates the condition number of the Information Matrix versus the input's center frequency. Figure 18(b) shows the plant's magnitude response curve with a shaded area superimposed indicating the input frequency band which yields the best (minimum) condition number. We see that the condition number is minimized by an input centered near the anti-resonant frequency (0.4 NFU). Thus, if the condition number of the Information Matrix is our metric of choice, this experiment suggests that we should excite the plant with inputs consisting of frequencies near a minimum in the plant's magnitude curve. Next, the trace of the estimation error covariance matrix (Figures 19(a) and (b)) clearly indicate that the sum of the error variances is minimized if the input is concentrated around the dip in the plant's frequency response.

This experiment tends to support one conclusion drawn in Chapter IV. When faced with output power constraints and white measurement noise, Equation (108) shows that the optimal input spectrum is shaped as the inverse of the plant. Thus, the optimal shaping filter tends to accentuate input frequencies near which the plant's magnitude is attenuated. Although this experiment's input sequence consisting of a sum of sinusoids is quite different from the sequence taken as the output of a shaping filter, we are encouraged by the agreement between the two approaches. That is, the optimal input frequency spectrum derived from maximizing entropy and the 'best' placement of sinusoids observed in this experiment both lead us to concentrate input energy in the same frequencies.

5.2.2 Bandpass Filters as Input Shaping Filters. The previous section gives us some agreement between the optimal shaping filter derived in Chapter IV and the best frequencies for sinusoidal inputs. Since this research is focused on finding
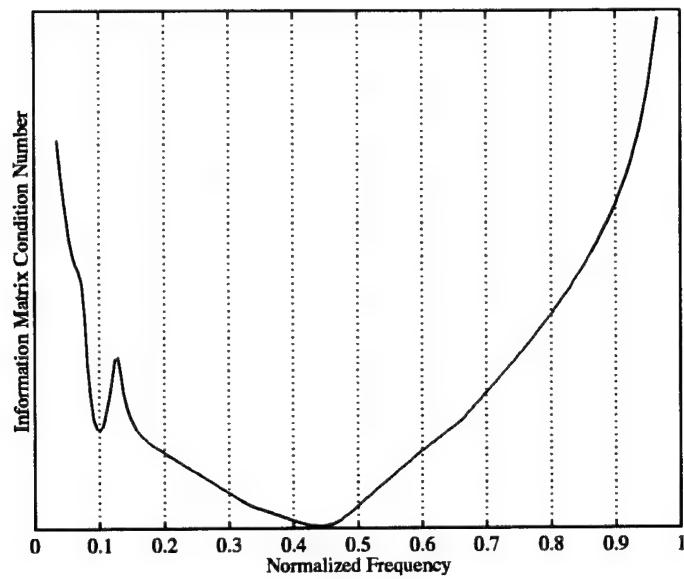
(a) Information Matrix Condition Number
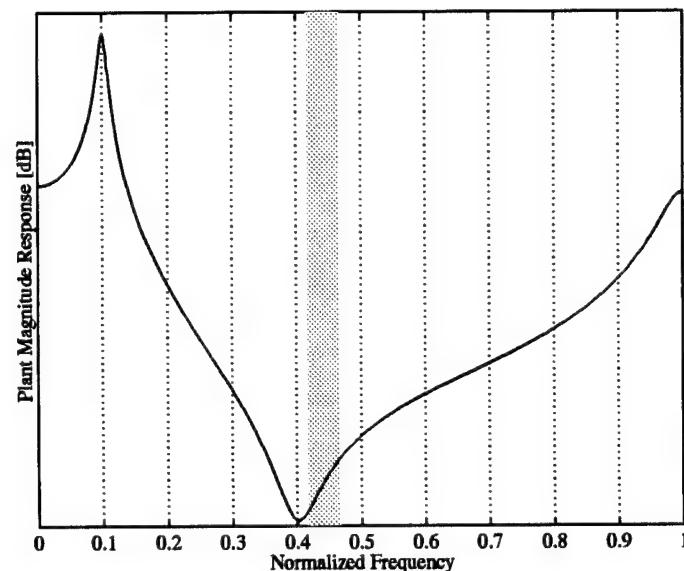
(b) Sinusoidal Band Yielding the Best Condition Number

Figure 18.   Information Matrix Condition Number: *(a) shows the condition number versus the center of the band of sinusoidal input. (b) shows the plant's magnitude response with the band of input frequencies which yields the best condition number.*
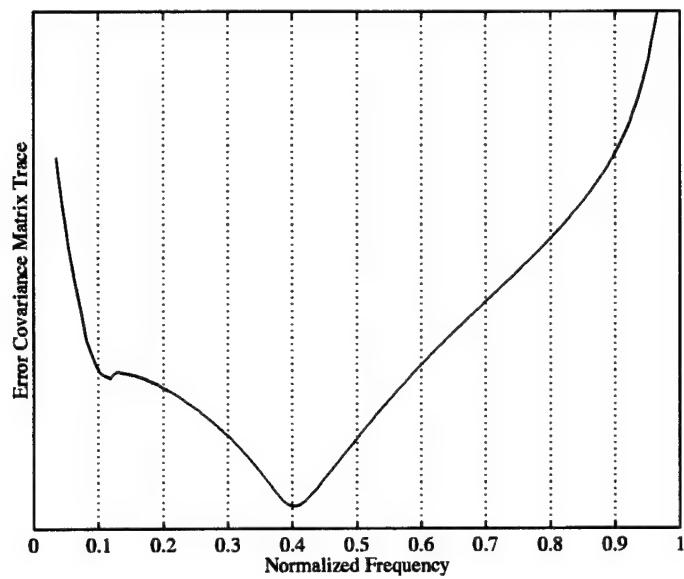
(a) Error Covariance Matrix Trace
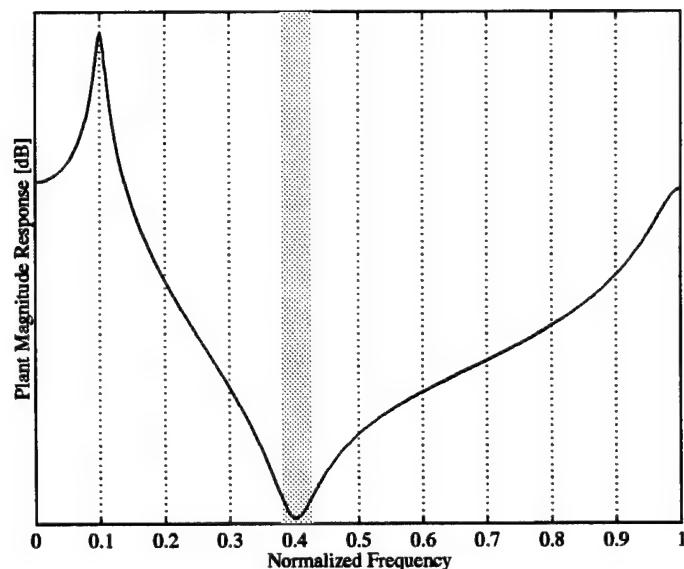
(b) Sinusoidal Band Yielding the Best Trace

Figure 19.   Error Covariance Matrix Trace: *(a) shows the trace of the error covariance matrix versus the center of the band of sinusoidal input. (b) shows the plant's magnitude response with the band of input frequencies which yields the best error covariance matrix trace.*

filters to be used to shape the input sequence, our next logical set of experiments involve the use of some sort of filter driven by white noise.

A bandpass input shaping filter is a good choice for our next experiment. Since a narrow bandpass filter can isolate a small set of frequencies, we can use several bandpass filters with different center frequencies to parallel the experiment we discussed in Section 5.2.1. As before, we wish to find the answer to the question: "What frequencies should be accentuated for optimal identification performance?" We attempt to form an answer by constructing input sequences as the outputs of different bandpass filters, forming estimates of the ID performance metrics.

The strategy employed for these experiments is similar to that used for the sinusoidal inputs. We form a bandpass filter with a specified pass-band center and bandwidth. The filter is excited with white noise and the resulting output is fed to the plant as input, after applying a gain to yield a predetermined average power on the output of the plant. We then use the plant's input and output sequences to build a regressor and calculate the ID performance metrics. Since we are dealing with pseudo-random sequences, we conduct the same experiment many times in a Monte Carlo fashion, averaging the calculated metrics over the number of experiments. After the desired quantities are stored, we move the center of the bandpass filter and begin the process over.

Again, Figure 17 helps to clarify this procedure. Each hatched region in this figure represents the ideal filter magnitude characteristic for one phase of this experiment. For example, we begin by designing a filter with a specified bandwidth and center frequency of ω_1. This filter is used to shape the input sequence. The resulting plant output and the filter's output are then used to construct the Information Matrix from which we calculate the metrics. We perform this procedure many times and store the mean of the metrics. We then move the filter's center frequency to ω_2 and re-accomplish another Monte Carlo set for the new input filter.
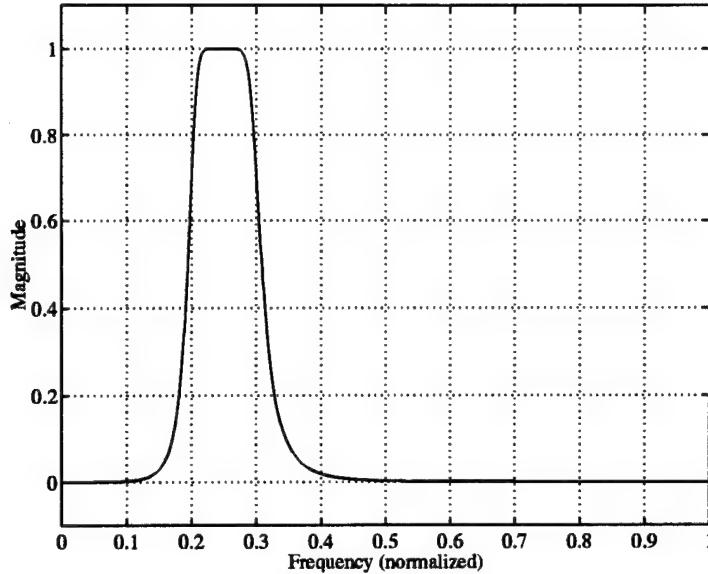
Figure 20. Magnitude Characteristic for a Typical Bandpass Filter

In order to facilitate easy comparisons, we use the same truth-model that we used for the sinusoidal experiment. This plant's Bode magnitude and phase plots can be seen in Figure 16. Each filter used in this experiment is an eighth-order Butterworth bandpass filter with a 3 dB bandwidth of 0.05 NFU. Our rationale for choosing this bandwidth is similar to that used in choosing the bandwidth in the sinusoidal experiment. We wish to isolate a small region of frequencies, but the bandwidth cannot be too small without incurring numerical difficulties. A typical bandpass filter's magnitude characteristic is shown in Figure 20.

In order to create a smooth plot, we quantize the frequency domain into 200 different band-center frequencies. Each filter is used in a Monte Carlo simulation consisting of 500 runs. Thus, we perform a total of 100,000 runs. Furthermore, each run consists of exciting the plant with a very long input sequence (10,000 points in time), from which we extract the Information Matrix based on a regressor matrix of length $q = 80$. We choose 10,000 points as an arbitrarily large time, allowing any transients to die out. The length of the regressor matrix (80 samples) represents a compromise between computation time and smoothness of the resulting plots.

98

We see the results of this experiment in Figures 21 and 22. Figure 21(a) shows the mean of the Information Matrix condition number versus the bandpass center frequency. The condition number is minimized by inputs with energy concentrated near the dip in the plant's magnitude response curve. Figure 21(b) shows the plant's magnitude versus frequency curve with the 'best' bandpass magnitude superimposed. Note that both curves are gain-modified so they can appear on the same plot for comparison. This figure clearly shows that the 'best' input (of the 200 tested) is band-limited to capture the plant's anti-resonant dip. Thus, input energy is concentrated most effectively in frequencies near those frequencies for which the plant attenuates the input. Figures 22(a) and (b) illustrate this point more strongly. These figures represent the trace of the error covariance matrix versus pass-band center frequency. As the curves show, minimizing this trace requires input energy centered on the dip in the plant's magnitude. Thus, this experiment also supports using inputs colored as the inverse of the plant as optimal.

5.2.3   Notch Filters as Input Shaping Filters.   The previous experiments allowed us to investigate optimal frequency- domain placement of input energy for the enhancement of ID performance. The experiments covered in this section are also concerned with the concentration of input energy, but here we attempt to infer the best place to *exclude* input energy. We accomplish this in a manner similar to that covered in Section 5.2.2. However, now the shaping filter is constructed as a notch filter, or an all-pass filter which blocks only a specified band. A typical notch filter magnitude characteristic is shown in Figure 23. This filter has a notch width of 0.1 NFU with a center frequency of 0.25 NFU.

Our procedure is similar to that used in the bandpass experiments. A notch filter is formed with a specified center and notch-width. The filter is fed white noise and the filter's output is used to excite the plant. Output and input measurements are then used to build a regressor matrix from which we calculate the same metrics we have seen in previous experiments. Again, each notch filter is used 500 times
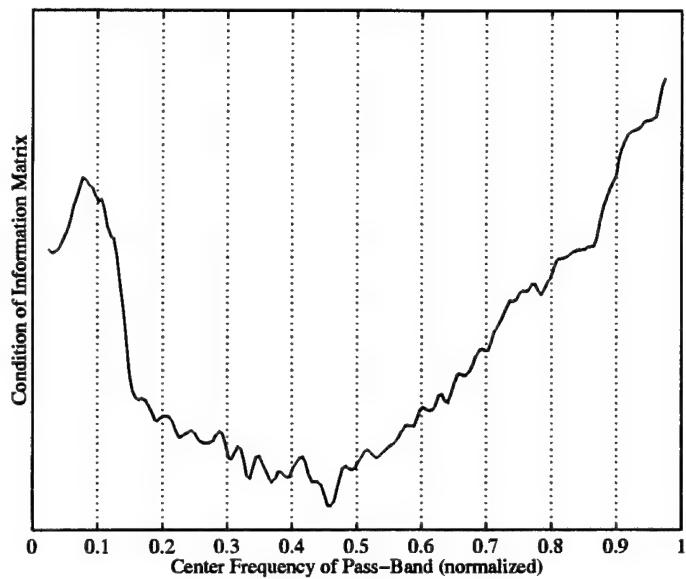
99

(a) Information Matrix Condition Number
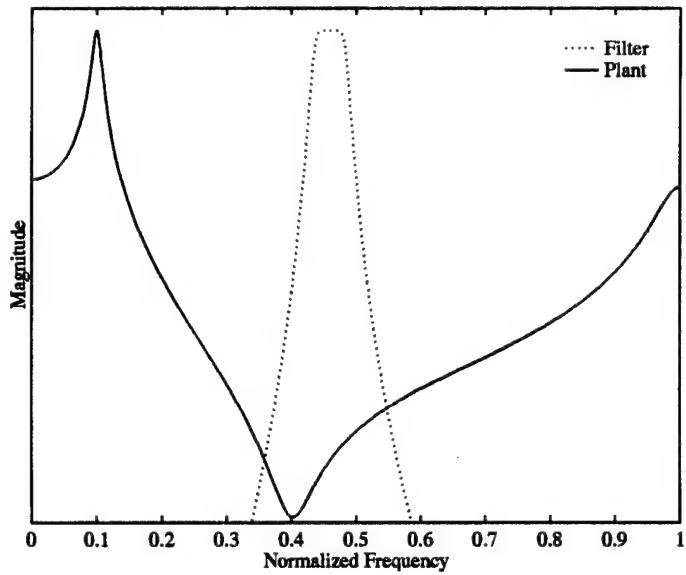
(b) Pass-Band Yielding the Best Condition Number

Figure 21.   Information Matrix Condition Number for the Plant Illustrated in Figure 16 Driven by Bandpass-Filtered Noise: *These plots show the condition number of the Information Matrix versus the center frequency of an input-coloring bandpass filter. (a) shows the condition number as a function of the center of the pass-band. (b) shows the plant's magnitude response (solid) and the bandpass filter magnitude characteristic (dotted) which yields the best condition number.*
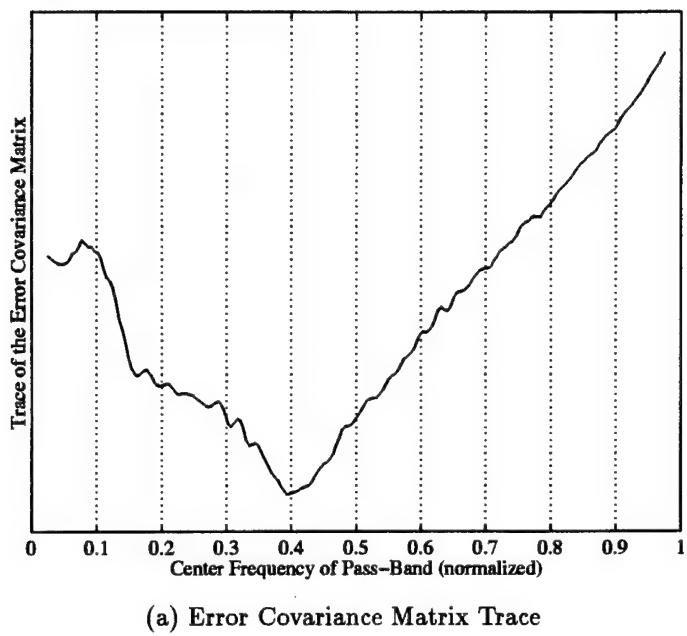
(a) Error Covariance Matrix Trace
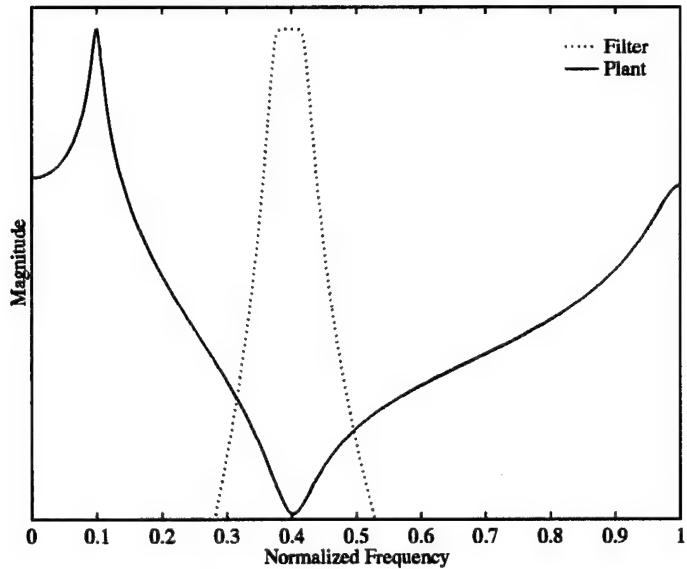
(b) Pass-Band Yielding the Best Trace

Figure 22.　Error Covariance Matrix Trace for the Plant Illustrated in Figure 16 Driven by Bandpass-Filtered Noise: *These plots show the trace of the error covariance matrix versus the center frequency of an input- coloring bandpass filter. (a) shows the trace as a function of the center of the pass-band. (b) shows the plant's magnitude response (solid) and the bandpass filter magnitude characteristic (dotted) which yields the best error covariance trace.*

101
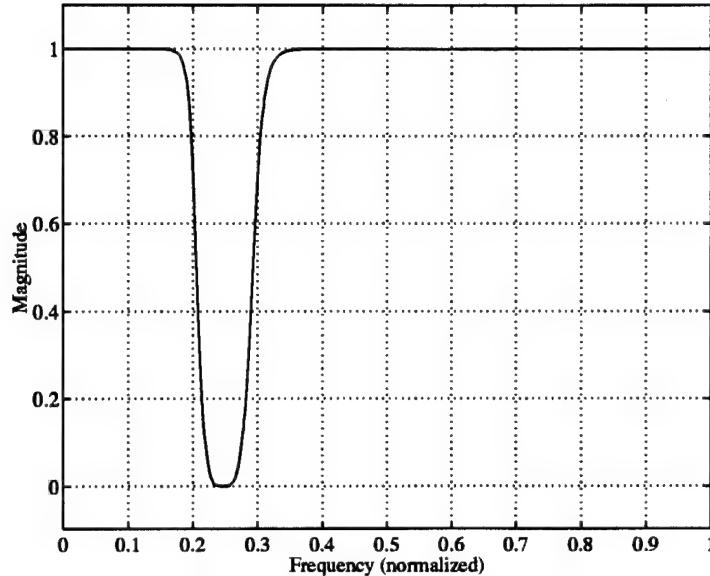
Figure 23. Magnitude Characteristic for a Typical Bandstop Filter

in a Monte Carlo analysis wherein the desired metrics are averaged over the number of runs. After conducting this procedure for each notch center frequency, the performance metrics are plotted versus the center frequencies. Following the same rationale outlined in Section 5.2.2, we set the notch-width to 0.05 NFU; we consider 200 notch-center frequencies equally spaced through the frequency domain; we use regressor-matrix lengths of 80 samples taken from the ends of 10,000 point runs.

This experiment further supports using the inverse of the plant as a shaping filter for enhanced identifiability. Figure 24 shows that the Information Matrix's condition number is minimized when the input's PSD excludes frequencies centered on the plant's resonant peak. Likewise, we minimize the estimation error covariance matrix's trace by excluding this same band of frequencies, as is shown in Figure 25. Therefore, out of the 200 different notch filters tested, the one which yields the highest potential for identifiability is the filter whose magnitude most closely approximates the inverse of the plant's magnitude.
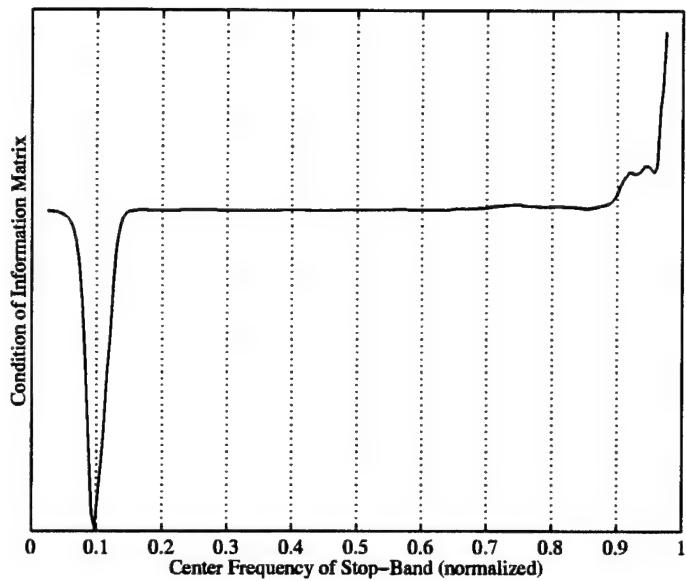
(a) Information Matrix Condition Number
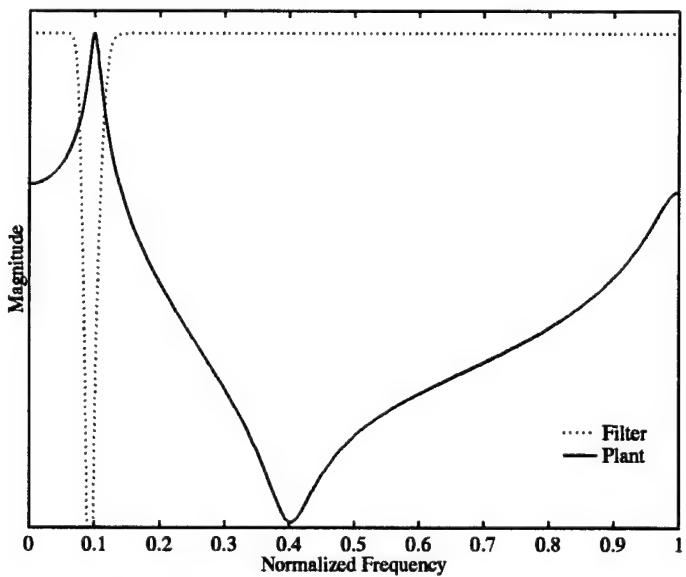
(b) Pass-Band Yielding the Best Condition Number

Figure 24.   Information Matrix Condition Number for the Plant Illustrated in Figure 16 Driven by Bandstop-Filtered Noise: *These plots show the condition number of the Information Matrix versus the center frequency of an input-coloring bandstop filter. (a) shows the condition number as a function of the center of the stop-band. (b) shows the plant's magnitude response (solid) and the bandstop filter magnitude characteristic (dotted) which yields the best condition number.*

103
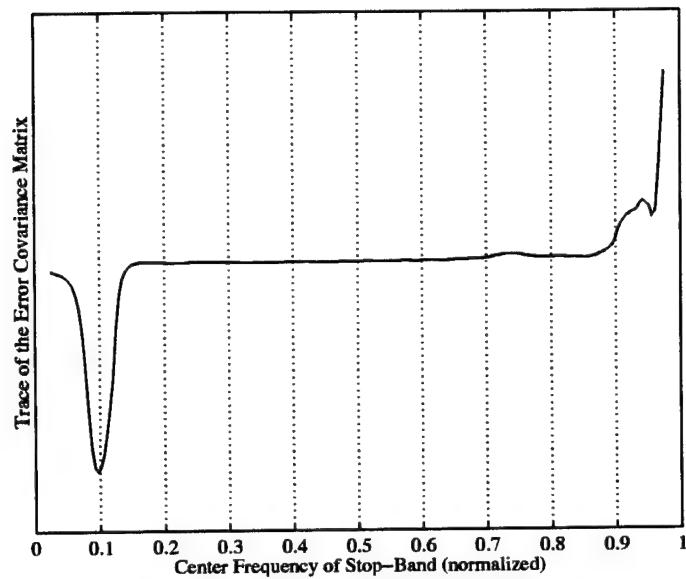
(a) Error Covariance Matrix Trace
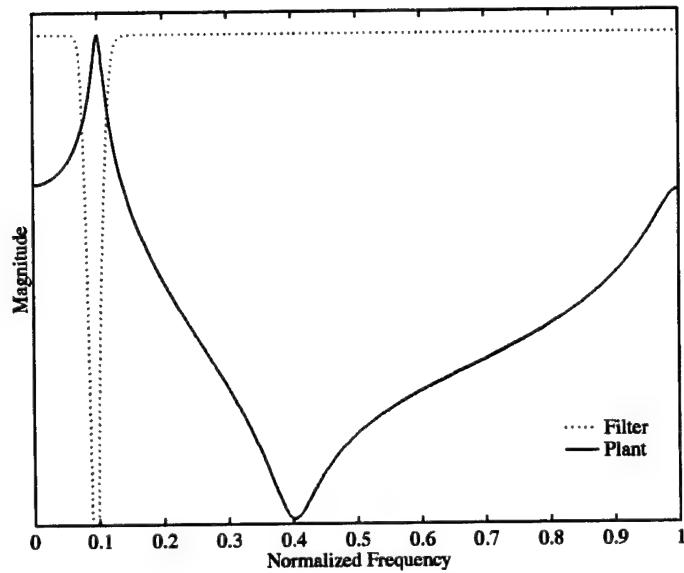
(b) Pass-Band Yielding the Best Trace

Figure 25. Error Covariance Matrix Trace for the Plant Illustrated in Figure 16 Driven by Bandpass-Filtered Noise: *These plots show the trace of the error covariance matrix versus the center frequency of an input- coloring bandpass filter. (a) shows the trace as a function of the center of the pass-band. (b) shows the plant's magnitude response (solid) and the bandpass filter magnitude characteristic (dotted) which yields the best error covariance trace.*

104

5.3 Parameter Estimation

We move now to the problem of estimating the parameters defining the mathematical model for a dynamical system. We saw in Chapter IV that the optimal input PSD for ID is inversely proportional to the plant's PSD when output power is constrained and the measurement noise is white. Furthermore, the experiments documented in previous sections of this chapter suggest an input PSD which approximates the inverse of the plant (at least around frequencies of resonance and anti-resonance).

The experiments documented in this section are designed to illustrate the superiority of the plant's inverse used as a shaping filter. Now, the problem of finding the optimal frequency shape of the input entails maximizing (or minimizing) a functional (i.e. functional optimization). Therefore, the domain of possible solutions has infinite dimensions, making the challenge of comparing all possible input PSD's clearly impossible to undertake.

Since we cannot explore all possible input frequency shapes, we will use straight white noise input as a baseline for comparison. This choice is arbitrary, but not without motivation. First, white noise is easy to implement; all we need is a random number generator and a gain. Second, white noise *seems* like a good choice because it contains all frequencies, ensuring that all modes of the plant are excited.

The methodology we employ in these experiments is straightforward Monte Carlo analysis. The plant is excited with the chosen input and the ID algorithm outlined in Appendix A is used to extract parameter estimates from the input sequence and a noise-corrupted measurement of the output. We repeat this procedure many times over many measurement noise realizations and compile statistics concerning the parameter estimates and the associated errors. Specifically, we run the estimation 100 times with measurement noise added to create a SNR of 40 dB. Note also that we adjust all input sequences to achieve an average output power of unity.
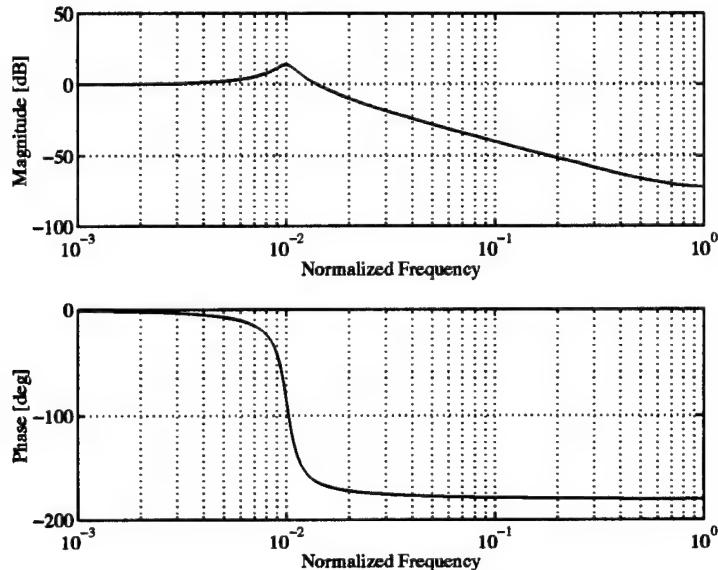
Figure 26. Bode Plot for the Plant to be Identified.

Thus, the measurement noise is simply a white Gaussian random sequence with a variance of 0.01.

5.3.1 Identification of an "Academic" Plant. In contrast to the models used in the previous experiments, this plant's poles are situated at much lower frequencies. We used high-frequency poles and zeros previously in order to facilitate plotting the effects near resonant frequencies. However, we wish now to illustrate the ID algorithm's ability to estimate the parameters of a model which represents a more realistic situation. That is, the sampling frequency is much higher than any dominant mode. Specifically, the plant is second order with poles situated at 0.01 NFU and a damping ratio of $\zeta = 0.1$. The transfer function is given by

$$\mathbf{G}(z) = \frac{b_1 z}{z^2 + a_1 z + a_2} = \frac{9.9371 \times 10^{-4} z}{z^2 - 1.9927z + 0.99371} \tag{124}$$

This model's Bode magnitude and phase plots are shown in Figure 26.

Once again, we are concentrating on the scenario in which output power is constrained and the measurement noise model is white. Thus, the optimal input PSD is proportional to the inverse of the plant's magnitude response squared.

106

Figures 27 through 32 give the results of this ID experiment. These plots clearly show us that the plant's inverse used as an input shaping filter yields markedly better identification compared to the use of white noise input. Figure 27 shows a good example. We see that the estimate bias on the a_1 parameter is reduced to an almost negligible level when the input is optimally colored, while white noise input yields a significant bias. Furthermore, we see that the RMS error for a_1 is reduced by about an order of magnitude by using the plant's inverse as an input shaping filter. This effect is continued for the rest of the parameters. Note especially the RMS error plots; RMS error captures both bias error and error variance.

5.3.2   Identification of an Air-to-Air Missile's Dynamics.   We now identify a realistic plant which represents a typical tail-controlled air-to-air missile. The transfer function presented here is taken from Brown's [34] M.S. thesis and related paper [35]. We consider the longitudinal acceleration versus pitch fin deflection. The continuous-time transfer function of the missile plant is

$$\mathbf{G}(s) = \frac{10.091(s^2 + 1.5580s - 774.18)}{s^3 + 42.442s^2 + 292.98s + 7812.5}$$

Transforming **G** into discrete-time form yields

$$\mathbf{G}(z) = \frac{10^{-2} \times (2.3956z^2 - 4.7936z + 2.3864)}{z^3 - 2.8975z^2 + 2.7970z - 0.89933}$$

where we have assumed a sampling rate of 400 Hz and a zero-order hold on the input. The Bode plots for this transfer function appear in Figure 33.

Figures 34 – 39 illustrate the results of this experiment. We first look at the errors committed in the autoregressive parameters. Figures 34 – 36 show us that optimally coloring the input yields estimates of the autoregressive parameters which are slightly better than those estimates calculated from white-noise input. However, the moving average (i.e. numerator) parameter estimates benefit dramatically from optimally coloring the input. Figure 37 shows that optimally colored input yields an improvement in the b_1 coefficient RMS error of about a full order of magnitude.
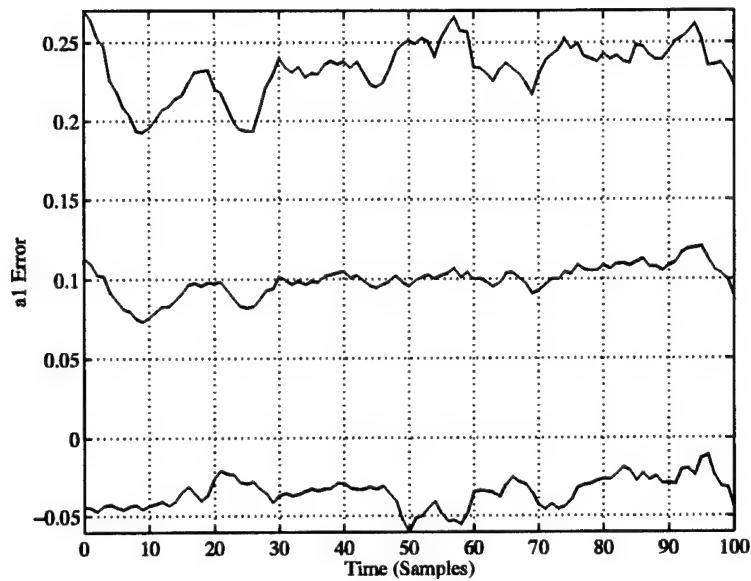
107

(a) White Noise Input
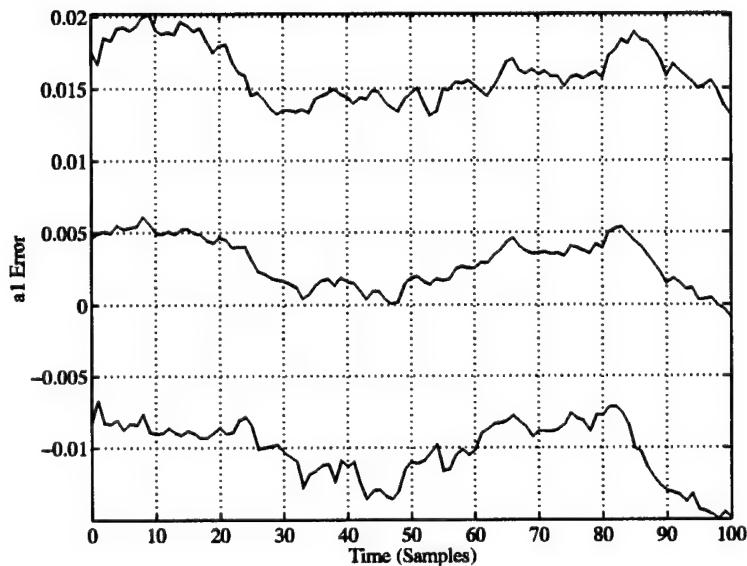
(b) Optimal Input for Output Power Constraints with White
Measurement Noise

Figure 27. a_1 Coefficient Error: *These plots compare the mean estimation error plus or minus one standard deviation committed on the a_1 coefficient. (a) shows the ID error for white noise input and (b) shows the ID error committed when the input is optimally colored.*

108
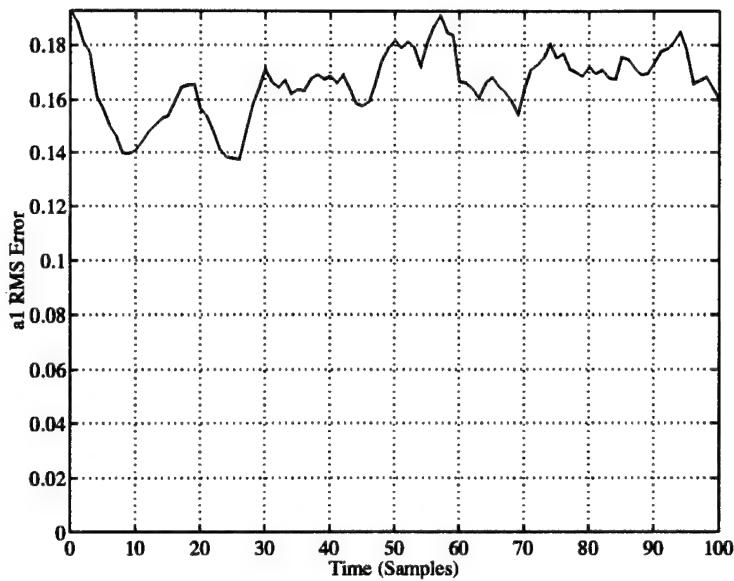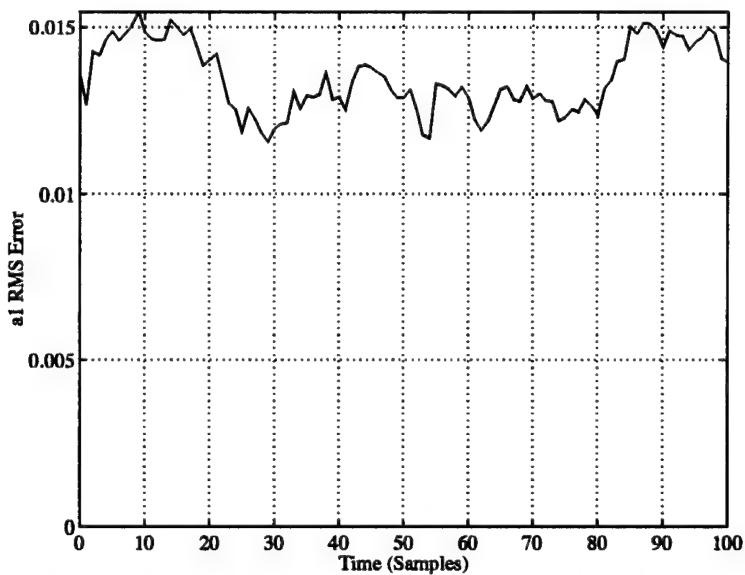
(a) White Noise Input

(b) Optimal Input for Output Power Constraints with White
Measurement Noise

Figure 28. a_1 Coefficient RMS Error: *These plots compare the RMS estimation error committed on the a_1 coefficient. (a) shows the ID RMS error for white noise input and (b) shows the ID RMS error committed when the input is optimally colored.*

109
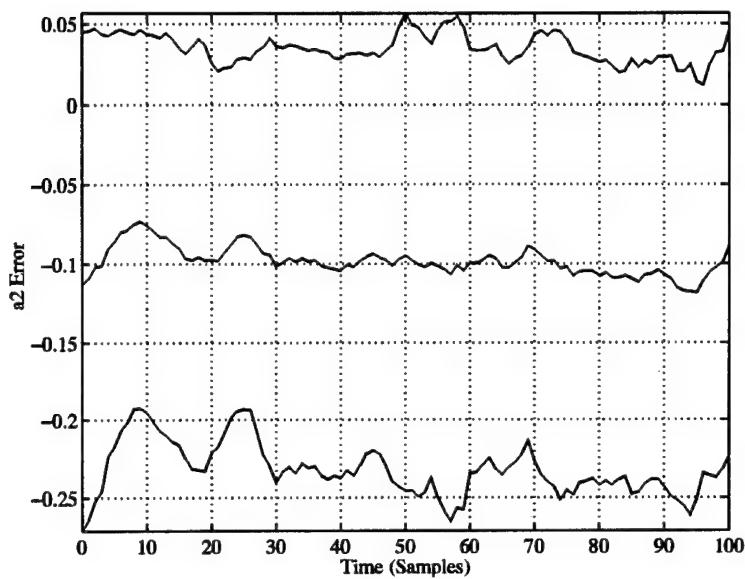
(a) White Noise Input
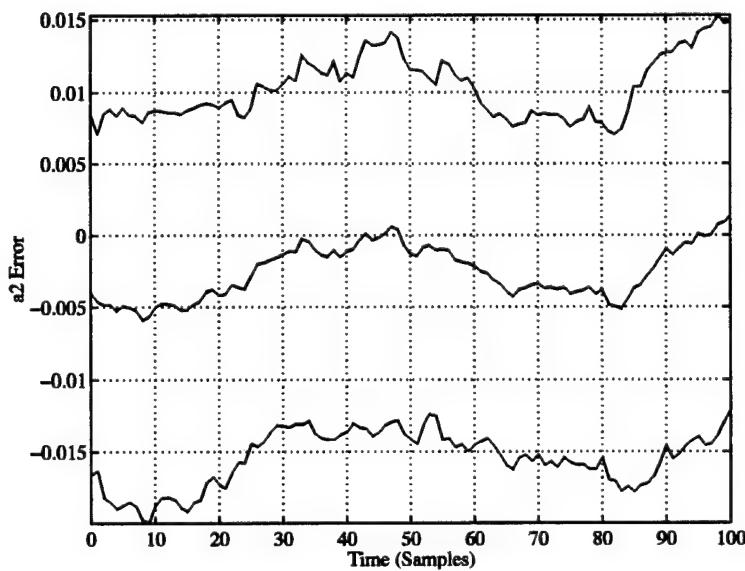
(b) Optimal Input for Output Power Constraints with White
Measurement Noise

Figure 29. a_2 Coefficient Error: *These plots compare the mean estimation error plus or minus one standard deviation committed on the a_2 coefficient. (a) shows the ID error for white noise input and (b) shows the ID error committed when the input is optimally colored.*

110
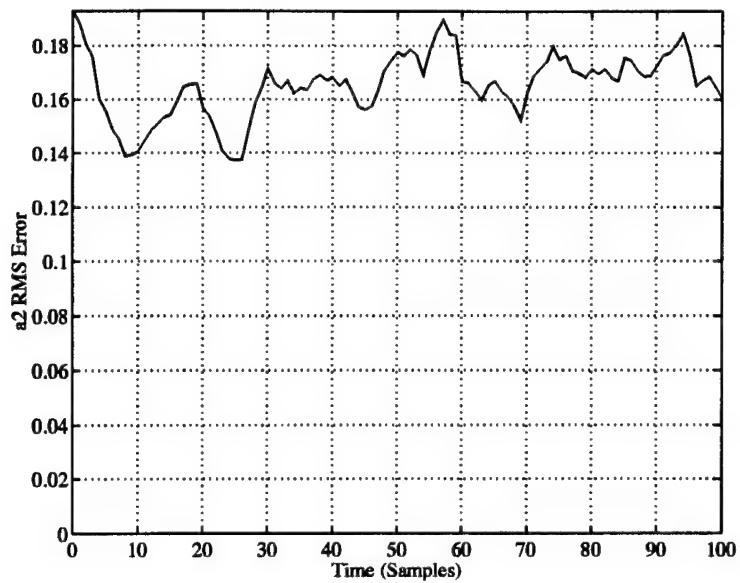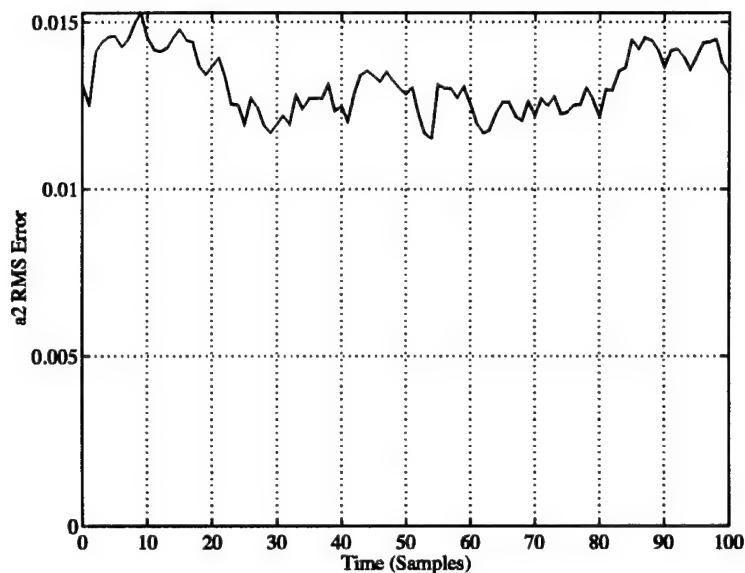
(a) White Noise Input

(b) Optimal Input for Output Power Constraints with White
Measurement Noise

Figure 30. a_2 Coefficient ID RMS Error: *These plots compare the RMS estimation
error committed on the a_2 coefficient. (a) shows the RMS error for white
noise input and (b) shows the ID RMS error committed when the input
is optimally colored.*

111
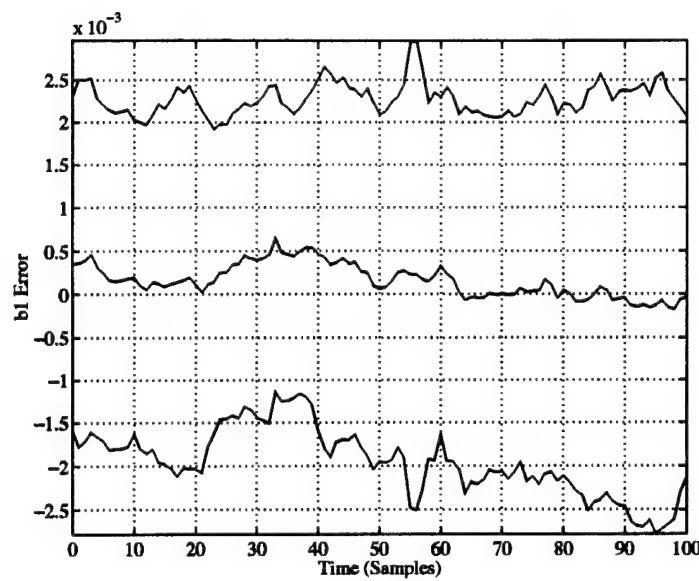
(a) White Noise Input
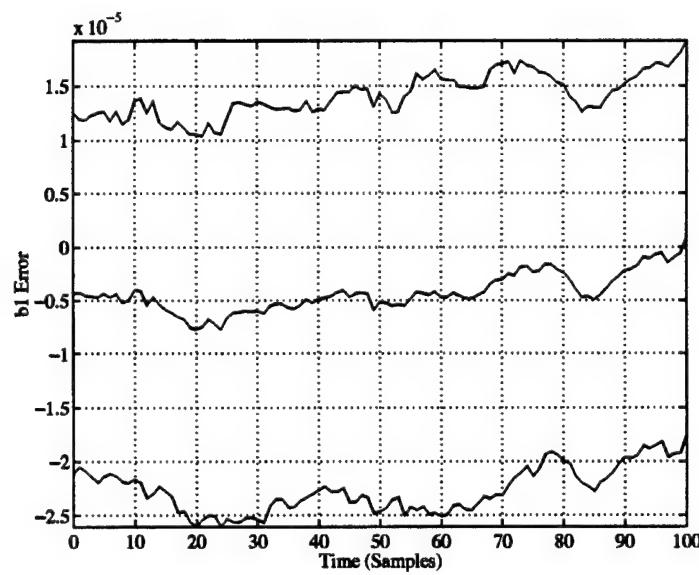
(b) Optimal Input for Output Power Constraints with White Measurement Noise

Figure 31. b_1 Coefficient Error: *These plots compare the mean estimation error plus or minus one standard deviation committed on the b_1 coefficient. (a) shows the ID error for white noise input and (b) shows the ID error committed when the input is optimally colored.*

112
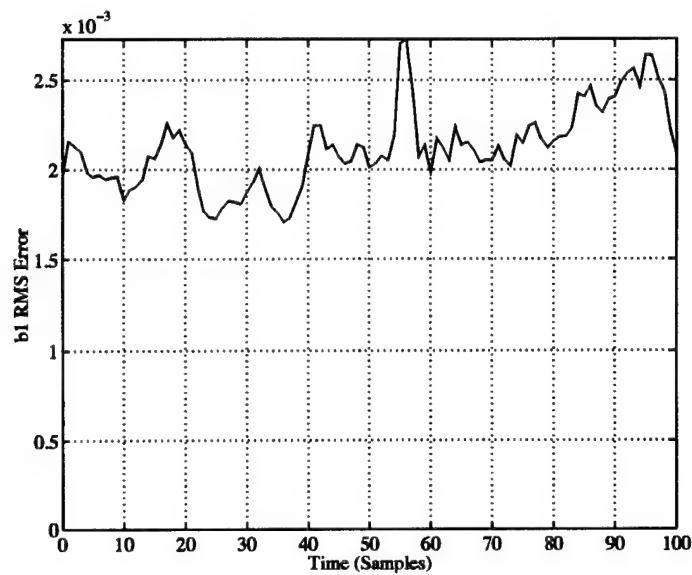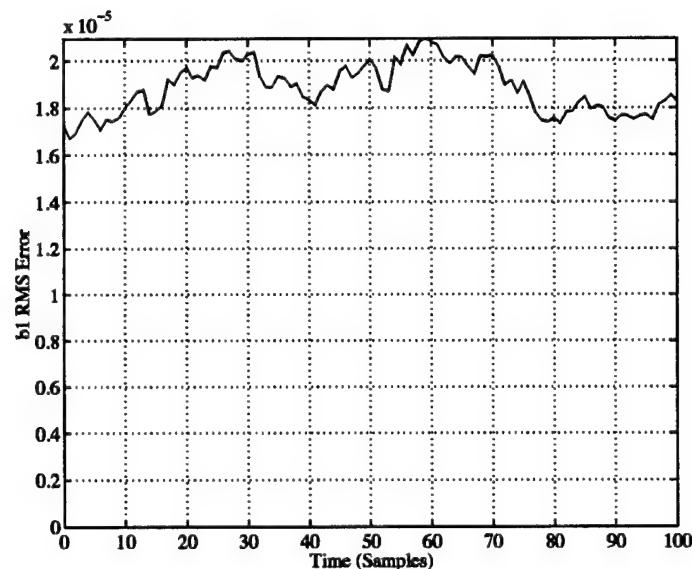
(a) White Noise Input

(b) Optimal Input for Output Power Constraints with
White Measurement Noise

Figure 32. b_1 Coefficient RMS Error: *These plots compare the ID RMS estimation error committed on the b_1 coefficient. (a) shows the RMS error for white noise input and (b) shows the ID RMS error committed when the input is optimally colored.*

113
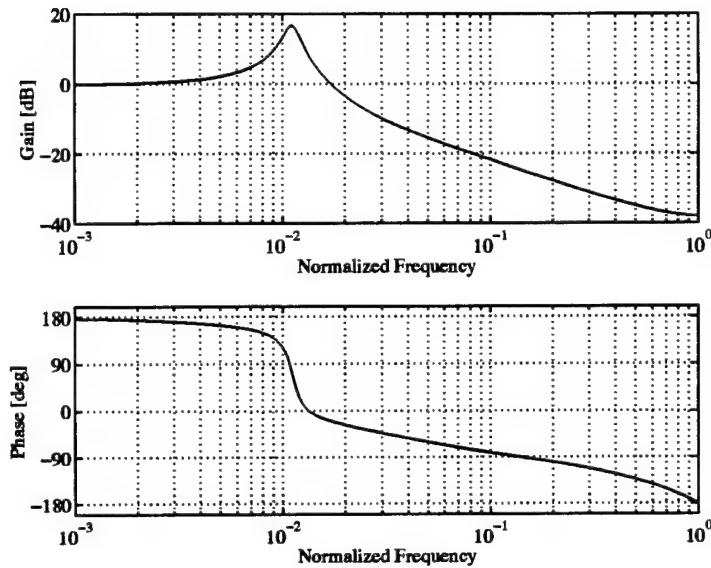
Figure 33.   Missile Bode Plots: *These Bode plots show the frequency response for the air-to-air missile's longitudinal acceleration due to pitch fin deflection.*

The RMS error on the b_2 and b_3 parameters is reduced by about half through the use of colored input versus white input.

5.4   Discussion

This chapter presented several experiments relating input frequency content to parameter identifiability. We looked at the distribution of sinusoidal inputs, inputs defined by the outputs of various bandpass filters, and notch filters used as input shaping filters. Finally, we saw the results of an identification experiment comparing white noise input to the use of the plant's inverse as an input shaping filter.

All these experiments support the optimality of using the plant's inverse as an input shaping filter. We saw that a sum of sinusoids is best grouped in frequency around a dip in the plant's magnitude response. Similarly, the 'best' bandpass filter is centered near an anti-resonant frequency. When the input is shaped by a notch filter, we can justify centering the notch near the plant's resonant frequency. Finally, actual parameter identification is dramatically improved by shaping the input's PSD
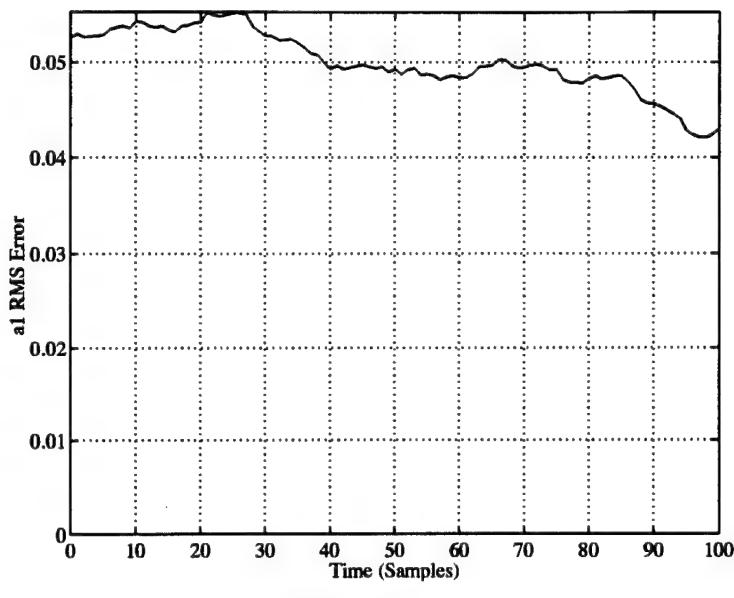
114
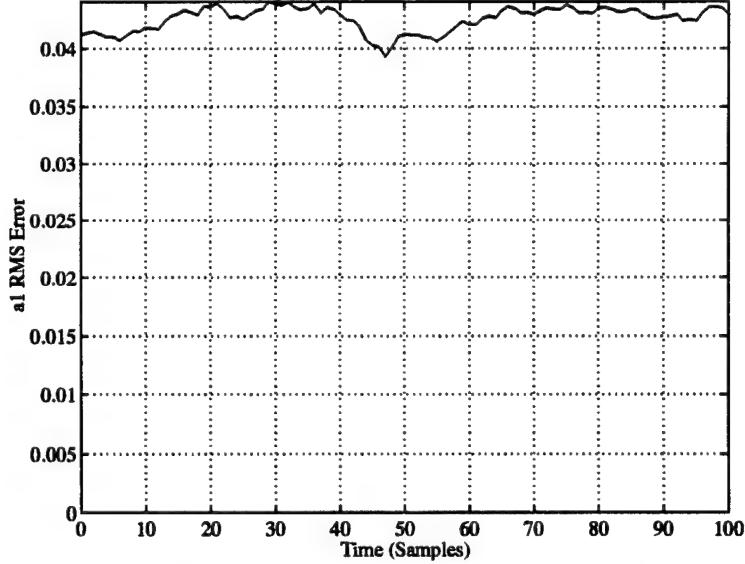
(a) White Noise Input

(b) Optimal Input for Output Power Constraints with White Measurement Noise

Figure 34. Missile ID – a_1 RMS Error: *These plots compare the RMS estimation error committed on the a_1 coefficient. (a) shows the ID RMS error for white noise input and (b) shows the ID RMS error committed when the input is optimally colored.*

115
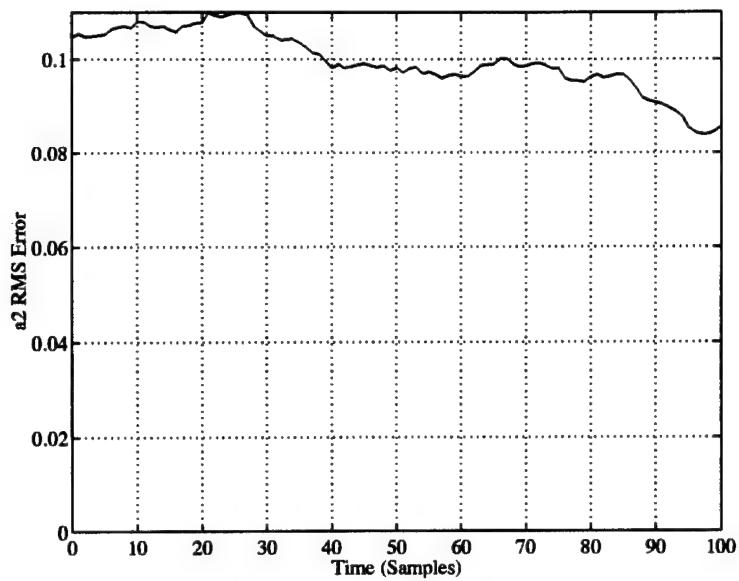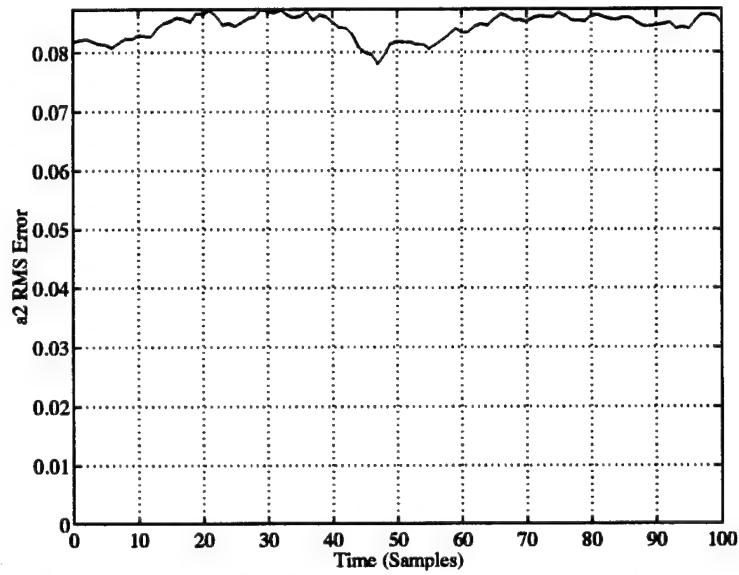
(a) White Noise Input

(b) Optimal Input for Output Power Constraints with White
Measurement Noise

Figure 35.   Missile ID – a_2 RMS Error: *These plots compare the RMS estimation error committed on the a_2 coefficient. (a) shows the ID RMS error for white noise input and (b) shows the ID RMS error committed when the input is optimally colored.*

116
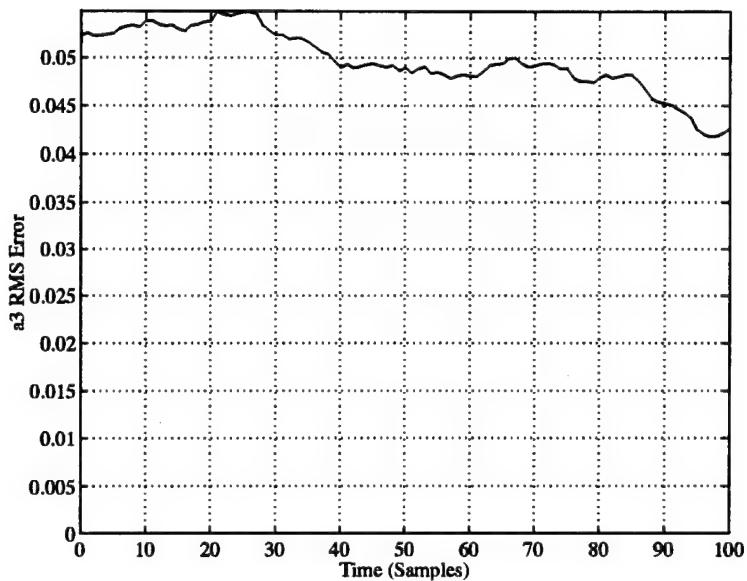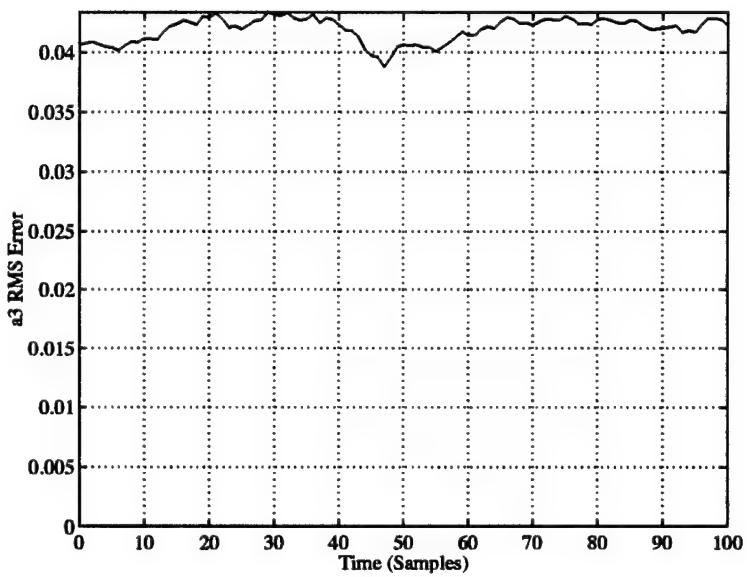
(a) White Noise Input

(b) Optimal Input for Output Power Constraints with White
Measurement Noise

Figure 36. Missile ID – a_3 RMS Error: *These plots compare the RMS estimation error committed on the a_3 coefficient. (a) shows the ID RMS error for white noise input and (b) shows the ID RMS error committed when the input is optimally colored.*

117
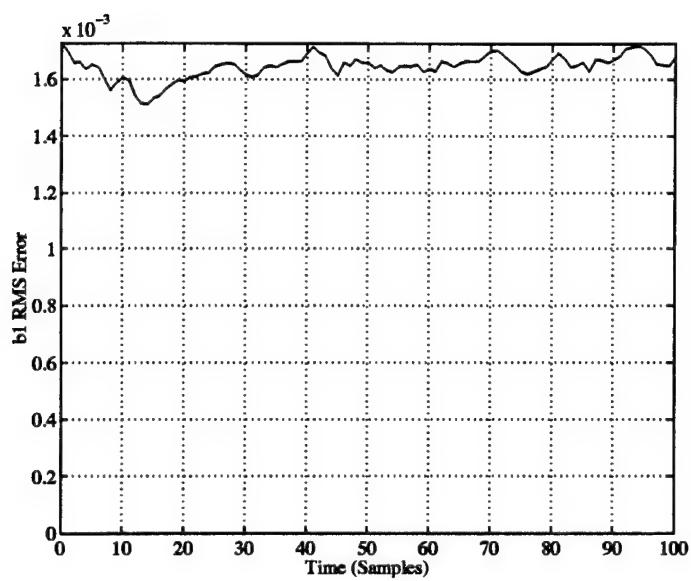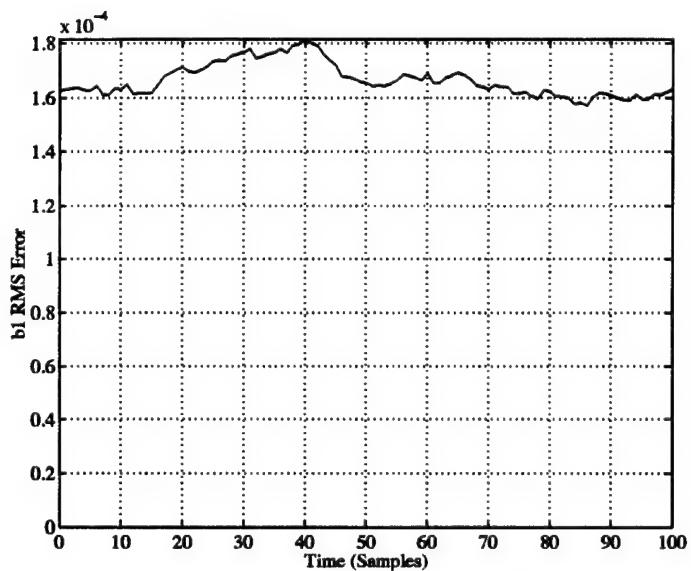
(a) White Noise Input

(b) Optimal Input for Output Power Constraints with White Measurement Noise

Figure 37.   Missile ID – b_1 RMS Error: *These plots compare the ID RMS estimation error committed on the b_1 coefficient. (a) shows the RMS error for white noise input and (b) shows the ID RMS error committed when the input is optimally colored.*

118
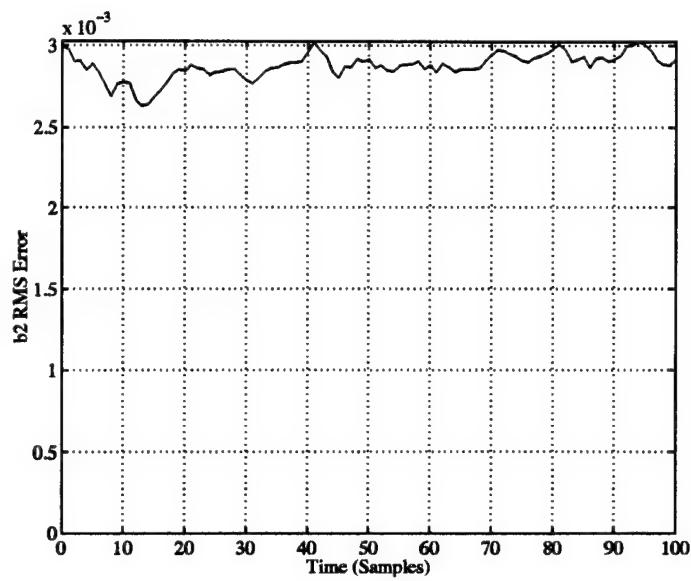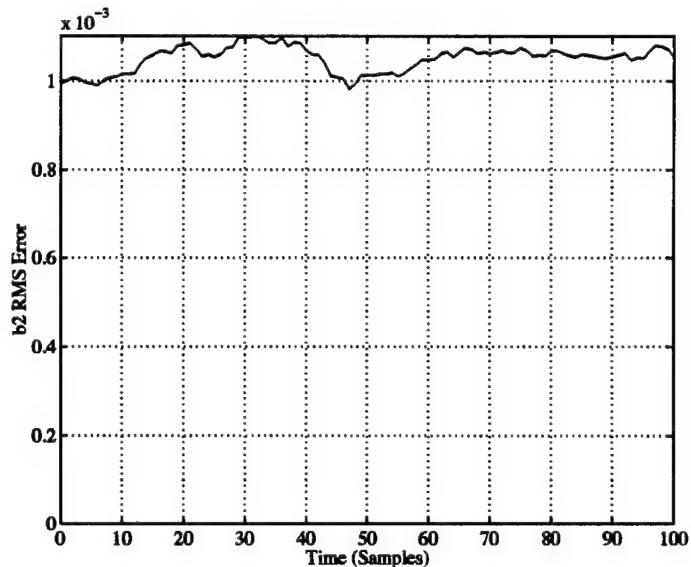
(a) White Noise Input

(b) Optimal Input for Output Power Constraints with White Measurement Noise

Figure 38.   Missile ID – b_2 RMS Error: *These plots compare the RMS estimation error committed on the b_2 coefficient. (a) shows the ID RMS error for white noise input and (b) shows the ID RMS error committed when the input is optimally colored.*

119
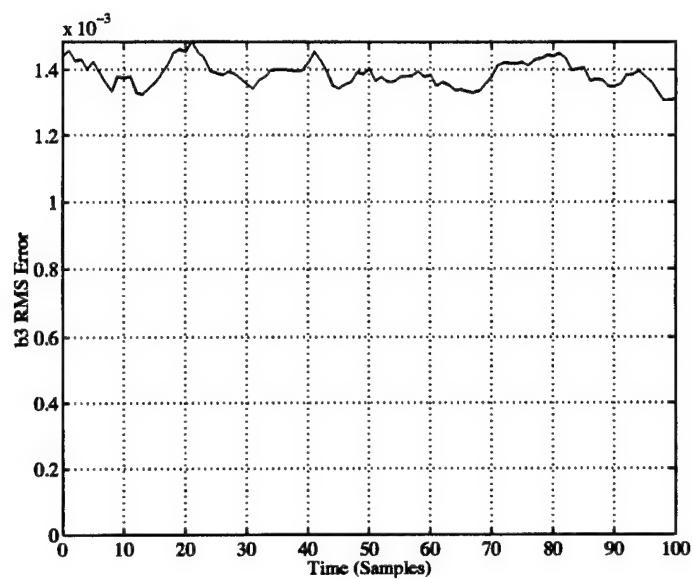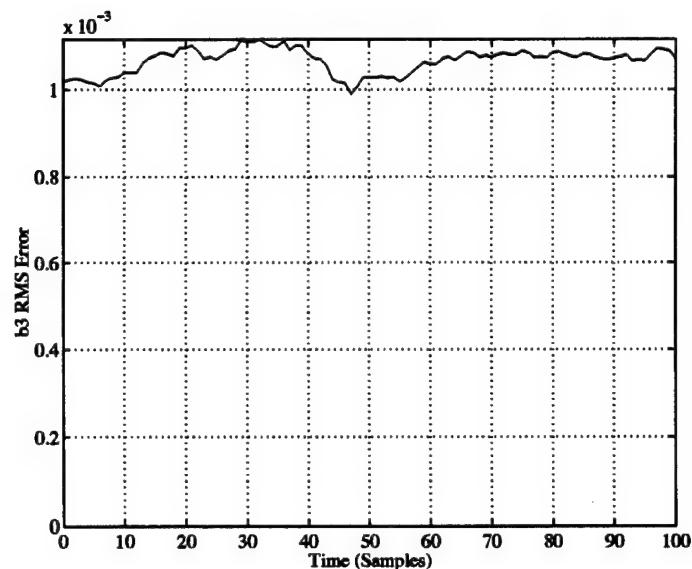
(a) White Noise Input

(b) Optimal Input for Output Power Constraints with
White Measurement Noise

Figure 39. Missile ID – b_3 RMS Error: *These plots compare the RMS estimation error committed on the b_3 coefficient. (a) shows the ID RMS error for white noise input and (b) shows the ID RMS error committed when the input is optimally colored.*

120

to be proportional to the inverse of the plant's PSD. Hence, we have confirmed the validity and usefulness of the theory developed in Chapter IV.

Furthermore, these experiments suggest a strong link between maximizing output entropy and minimizing estimation error. We derived the optimal input such that it provides the ID algorithm with as much information as possible. The experiments we conducted for this chapter clearly demonstrate that increasing the information content of the measured signals can reduce parameter estimation error.

We now have a tested technique for generating inputs which are optimal in an information theoretic sense. Furthermore, we have seen that these optimal inputs facilitate parameter estimates with higher accuracy (compared to results where white noise is used as an input). However, we may still wish to answer the question, "How close must the parameter estimates be?" Chapter VI provides a partial answer via some theory and experimental evidence.

VI. Limits of System Identification

Chapter V documented experiments which clearly show the utility of the theory developed in Chapter IV. Furthermore, we saw the results of experiments which suggest (through different metrics) that the optimal shaping filter has a PSD equal to the inverse of the plant's PSD, which corroborates the results of Chapter IV. In this chapter, we now concern ourselves with the practical limits of System Identification. In other words, we wish to explore the complexity of systems which can be successfully identified under the best possible conditions, and such that a level of parameter error which is acceptable is obtained. Furthermore, we will investigate the effects of Signal to Noise Ratio on ID accuracy.

6.1 Introduction

The role of System Identification (ID) is to generate models of the plant for the purpose of off-line or on-line control design (e.g., as required in indirect adaptive control). System Identification is a data-driven procedure where system inputs and noise-corrupted output measurements are used to estimate the parameters defining a dynamic system. In the case of linear, time-invariant systems, we are faced with identifying the coefficients of polynomials making up a rational transfer function. When applied to control design using frequency-domain methods, the control engineer is probably most interested in the poles and zeros of the dynamical system. Hence, one should be concerned with the accuracy of the roots (viz., the poles and zeros) and the frequency response of the identified system, rather than the coefficients of its transfer function. Therefore, one objective of this chapter is to illustrate the sensitivity of a polynomial's roots and the resulting transfer function's magnitude response with respect to identification errors in the coefficients; said errors are commensurate with ID accuracy.

The sensitivity of the roots of high-order polynomials with respect to the polynomial's coefficients is well known to Applied Mathematicians [5, 73]. Unfortunately, in engineering circles the belief sometimes persists that the very use of high-order models is conducive to higher accuracy. The above mentioned sensitivity problem is easily recognized upon performing Matlab® experiments using the *roots* and *poly* commands. Now, the source of error encountered by numerical mathematicians is numeric, and is, therefore, of a very low level. In contrast, the controls engineer must contend with realistic measurement noise, which greatly exacerbates the sensitivity problem.

This chapter is organized as follows. Section 6.2 states and explores the polynomial roots' sensitivity problem. Section 6.3 presents experimental results which illustrate the sensitivity of polynomial roots with respect to equal errors in each of the polynomial coefficients. Section 6.4 gives the effects of coefficient errors arising from realistic identification experiments which include measurement noise. Section 6.5 presents some interesting experiments that lead to the paradoxical situation in which a larger Signal to Noise Ratio can lead to higher estimation errors. Finally, Section 6.6 offers some conclusions.

6.2 Sensitivity of Roots to Coefficient Errors

Theoretically, the roots of a polynomial depend continuously on its coefficients. At the same time, the roots of a polynomial can be very sensitive to small errors in the coefficients of the polynomial. The following discussion, adapted from Ralston [73] (see also [5]), gives some insight into the problem.

Consider a polynomial of the form

$$f(z) = \sum_{i=0}^{n} a_i z^i \tag{125}$$

123

Perturb f by introducing errors in the polynomial's coefficients, viz., obtain the polynomial

$$
\begin{aligned}
g(z) &= \sum_{i=0}^{n}(a_i + \delta_i)z^i \\
&= f(z) + \sum_{i=0}^{n}\delta_i z^i
\end{aligned}
\tag{126}
$$

Now, if z_o is a root of f and $z_o + \epsilon_o$ is a root of the perturbed polynomial g, then

$$
\begin{aligned}
g(z_o + \epsilon_o) &= f(z_o + \epsilon_o) + \sum_{i=0}^{n}\delta_i(z_o + \epsilon_o)^i = 0 \\
&\approx \epsilon_o f'(z_o) + \sum_{i=0}^{n}\delta_i z_o^i
\end{aligned}
\tag{127}
$$

where

$$
f'(z_o) \triangleq \left.\frac{\partial f}{\partial z}\right|_{z=z_o}
$$

and we have assumed that δ_i and ϵ_o are 'small,' allowing us to ignore products of the errors.

Equation (127) gives an estimate for the error in the root:

$$
|\epsilon_o| \approx \frac{\left|\sum_{i=0}^{n}\delta_i z_o^i\right|}{|f'(z_o)|}
\tag{128}
$$

The utility of Equation (128) is limited when $f'(z_o)$ is small, as is the case when z_o is a repeated root. Of course, a very small $f'(z_o)$ implies that our initial assumption of small errors is incorrect, in which case Equation (128) provides a poor estimate of the error. Equation (128) also warns us that in the case of 'high-stiffness', where the roots of the polynomial are located in widely separated clusters, accurate roots will be difficult to resolve.

We can also experience problems with this estimate even when $f'(z_o)$ is large. Again, Ralston [73] gives the following example.

Consider a polynomial with roots $\{-1, -2, \ldots, -20\}$

$$
f(z) = (z+1)(z+2)\cdots(z+20)
\tag{129}
$$

124

If $z_o = -20$, then $|f'(z_o)| = 19!$ (certainly not small). Assuming a very small error in a_{19}, say $\delta_{19} = 10^{-7}$, then Equation (128) gives the error estimate

$$|\epsilon_o| \approx \frac{10^{-7}20^{19}}{19!} \approx 4.3 \tag{130}$$

which is not small relative to $z_o = -20$. Thus, we cannot trust the estimate of the error in this root due to an error in the z^{19} coefficient. However, by contrapositive argument, we *can* predict significant errors in the roots arising from the seemingly small coefficient error. A polynomial with this property is said to be ill-conditioned.

The previous discussion illustrates the effect of polynomial ill-conditioning. Obviously, an ill-conditioned polynomial may be factored accurately only if the coefficients are given to a high degree of precision. Unfortunately, high degree polynomials are usually ill-conditioned. In fact, "it is generally true that the solution of *high-degree* polynomial equations *requires* the use of multiple-precision floating-point arithmetic in order to achieve high accuracy" [73].

Now, Ralston [73] (see also [5]) is referring to miniscule errors committed by quantizing real numbers into floating-point representations required by digital computers. Indeed, such digital quantization noise is very small. Double precision allows about 17 decimal digits of precision, or approximately a 340 dB signal to noise ratio (SNR). Indeed, a good example of the requirement for very high numerical precision in control design is given by the frequency-domain Quantitative Feedback Theory CAD package described in [77, 78]. The user-defined precision capability of Mathematica® was used to achieve as much as 100 digits of precision (which represents a huge 2000 dB SNR!) in order to combat the significant deleterious effects of very small quantization noise. In contrast, System Identification relies on noisy measurements, and hence is operating on much 'dirtier' data; a typical SNR for ID in an engineering environment is closer to 40 – 60 dB. Thus, much of the accomplishments of Numerical Analysis is not applicable to ID work, since errors associated

with the identification of transfer function poles and zeros are driven primarily by large measurement noise.

Hence, we have seen how even the relatively sterilized environment of the digital computer, with errors created only by quantizing perfect data, can have marked effects on the accuracy of the roots of polynomials. Next, we investigate the sensitivity of roots of polynomials, where coefficients are obtained from ID experiments using data corrupted by 'real-world' levels of noise.

6.3 Experiments Using Equal Coefficient Errors

Consider a very simple seventh-order, all-pole plant represented by the transfer function

$$G(s) = \frac{8!}{(s+2)(s+3)\cdots(s+7)(s+8)} \tag{131}$$

Indeed, the plant's characteristic equation is fairly benign, for its stiffness (ratio of the highest to the lowest root [73]) is rather limited, and no double roots are present. We assume the ID scheme can identify the coefficients of the transfer function to within some bound, then construct polynomials consisting of the nominal seven coefficient values ± the error. In this case of a seventh-order polynomial, Matlab® is used to build $3^7 = 2187$ different polynomials, corresponding to every combination of the nominal coefficient values ± error. As each polynomial is constructed, Matlab® calculates the poles and the frequency response of the transfer function defined in Equation (131).

First, we set the relative error to ±1% for all coefficients. The scatter diagram in Figure 40(a) shows all possible roots (poles) for the 2187 different polynomials. The nominal roots (poles) are highlighted by an asterisk while the roots (poles) of the perturbed polynomials are depicted by dots. Notice how very small errors in the coefficients translate into large excursions in the placement of the roots. In particular, Figure 40(b) shows the *dominant* roots for each of the 2187 polynomials. These poles,
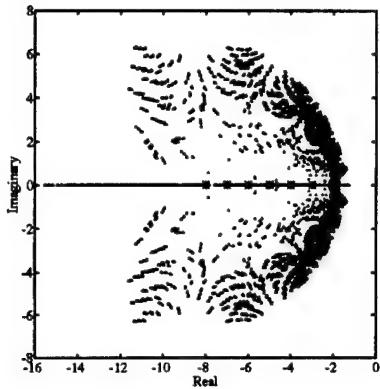
(a) Roots
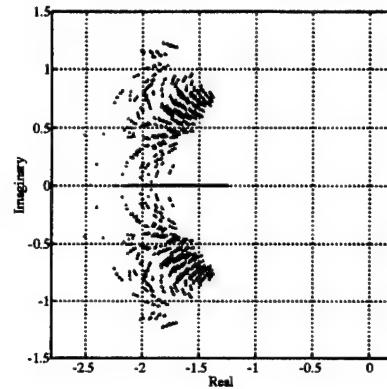
(b) Dominant Roots
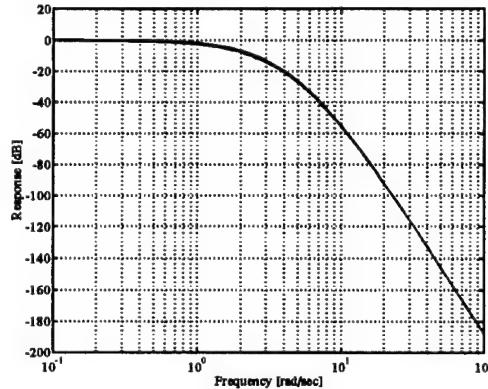
(c) Magnitude Response
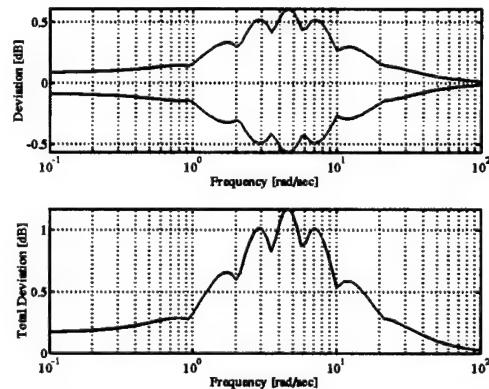
(d) Magnitude Response Deviation

Figure 40.   Coefficient Sensitivity (Seventh-Order Polynomial): *These plots show the sensitivity to 1% errors in the coefficients of a seventh-order polynomial. (a) and (b) show all roots and dominant roots, respectively. (c) shows the magnitude response for the nominal and perturbed characteristic polynomials. (d) shows the worst-case deviations from nominal. The nominal transfer function is given by* $G(s) = \frac{8!}{(s+2)(s+3)\cdots(s+8)}$.

which determine the time-domain response of the dynamical system, drift far enough to have a damping ratio of approximately 0.8. On the more encouraging side, Figures 40(c) and (d) show that the magnitude response of the perturbed transfer function deviates from the nominal by only about 0.6 dB, in the worst case.

The next Matlab® experiment is performed identically to the first, but here the coefficient errors are bounded by a seemingly acceptable level of accuracy of $\pm 10\%$. The fingerprint-like plots shown in Figures 41(a) and (b) illustrate that the roots (poles) now deviate markedly from the nominal. In fact, the dominant roots can drift far enough to exhibit a natural frequency of about $7\frac{rad}{sec}$ with a damping ratio as low as about 0.2. Clearly, this difference in root (pole) locations would have a significant effect on closed-loop stability. The effect of the drifting coefficients is further illustrated by the magnitude response curves shown in Figures 41(c) and (d). The 'small' 10% errors in the coefficients yield a transfer function error envelope which expands to about 14 dB wide, at relatively *low* frequencies. Hence, errors in the coefficients of the transfer function's polynomials manifest themselves as structured uncertainty.

6.4 Coefficient Errors

The previous section demonstrated the sensitivity of polynomial root (pole) placement to errors in the coefficients defining the transfer function. The next question is, "How much error do we reasonably expect in the identification of the coefficients?"

Several ID experiments generate a partial answer. These experiments are performed with different transfer functions, each based on a continuous-time system with poles placed according to Table 7. Each transfer function is converted into its discrete-time analog using a sampling time of $\frac{\pi}{32}$ (i.e. four times the fastest natural frequency). Next, the coefficients making up the Z-plane transfer functions' denominators are estimated via weighted least-squares Linear Regression, with the weighting matrix given by the inverse of the covariance of the 'equation-error' vector, as discussed in Chapter III.
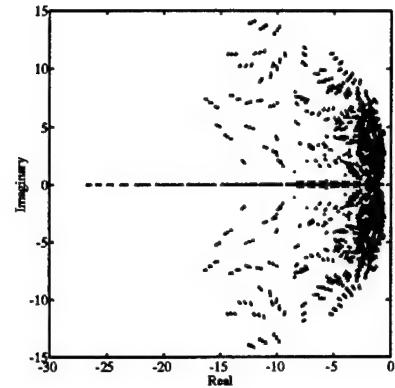
(a) Roots
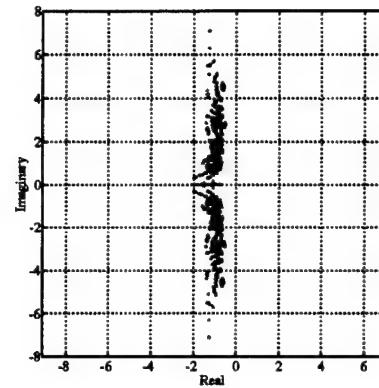
(b) Dominant Roots
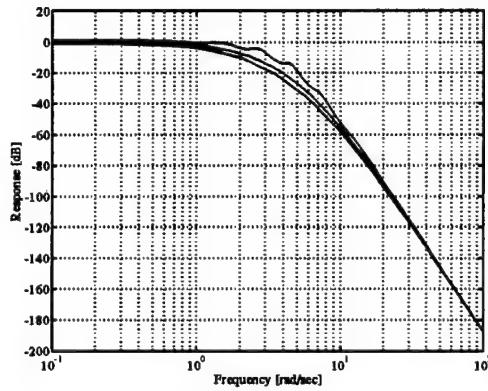
(c) Magnitude Response
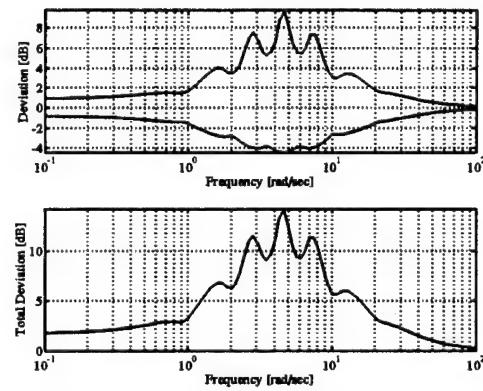
(d) Magnitude Response Deviation

Figure 41.   Coefficient Sensitivity (Seventh-Order Polynomial): *These plots show the sensitivity to 10% errors in the coefficients of a seventh-order polynomial. (a) and (b) show all roots and dominant roots, respectively. (c) shows the magnitude response for the nominal and perturbed characteristic polynomials. (d) shows the worst-case deviations from nominal. The nominal transfer function is given by* $G(s) = \frac{8!}{(s+2)(s+3)\cdots(s+8)}$.

6.4.1   Identification Using Weighted-Least-Squares.   The regressions for the experiments documented in this section are performed on a window of input/output pairs generated by exciting the 'plant' with noise colored by a shaping filter which approximates the inverse of the plant, motivated by theory developed in Chapter IV. The outputs are corrupted by Gaussian white noise designed to produce a SNR of

129

Table 7 Roots and RMS Coefficient Estimation Errors

Order	Roots	Coefficient Errors (% of Nominal)						
		a_1	a_2	a_3	a_4	a_5	a_6	a_7
2	-2,-8	0.07	0.2					
3	-2,-3,-8	0.11	0.32	0.75				
4	-2,-3,-4,-8	0.28	0.75	1.5	2.8			
5	-2,-3,-4,-5,-8	0.61	1.6	3.2	5.8	10		
6	-2,-3,-4,-5,-6,-8	1.1	2.9	5.8	11	19	32	
7	-2,-3,-4,-5,-6,-7,-8	1.6	4.3	8.7	16	29	52	92

100 dB – i.e. very 'clean' measurements are used. In each case, the regressor is constructed such that its length is $2m^2$ where m is the number of unknown coefficients. This regressor is used in a Weighted Least Squares scheme to estimate the parameters. The regression is performed 1000 times in a Monte Carlo fashion. The results are summarized in Table 7, where

$$G(z) = \frac{K}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_7 z^{-7}}$$

We see that as the model-order increases, the coefficient error increases dramatically. In particular, notice that the errors make a striking jump as the order of the plant increases from sixth to seventh-order. Furthermore, higher-indexed coefficients are harder to estimate.

For further insight, Matlab® is used to produce plots of the root locations and frequency response for each of the identified plants outlined in Table 7. Each experiment yields three plots. First, the root locations are illustrated – both actual roots and the roots of the identified system. Next, Bode magnitude and phase plots show the worst-case errors over the entire 1000 run Monte Carlo set. The worst-case envelopes are defined on a point-by-point basis throughout the frequency domain. Finally, the RMS errors in both magnitude and phase are presented as a function of frequency. Note that all frequencies are normalized such that unity corresponds to the Nyquist frequency.

130
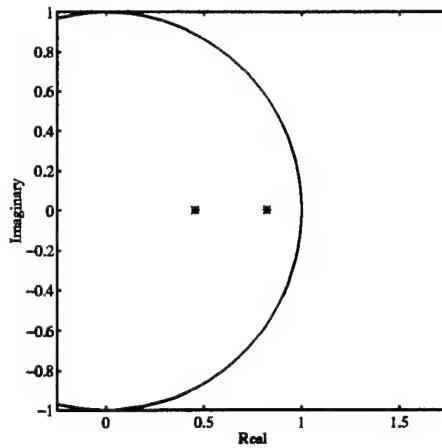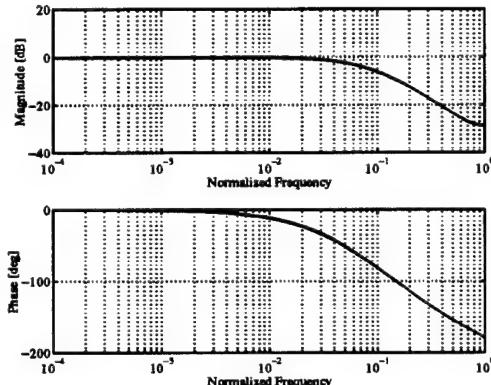
(a) Root Locations


(b) Worst-Case Envelopes
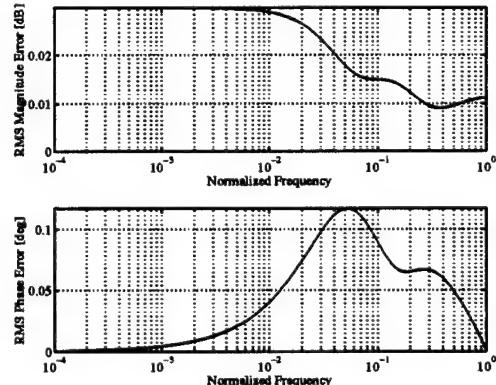
(c) RMS Errors

Figure 42.   Identification Errors (Second-Order System): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

The first experiment in this series involves the identification of a simple second-order plant with real roots. As Figure 42 shows, the identification is quite successful. We see very little error in root locations and magnitude and phase response. In fact, the dots (·) indicating the identified root locations are obscured by the asterisks (*) which represent the true root locations.

131

As we increase model-order from third through sixth-order, we see a gradual degradation in the identification. Figure 43(a) shows the migration of the root locations for the the third-order plant experiment. Although the roots are misidentified slightly, Figures 43(b) and (c) show that the ID error hardly manifests itself in the response curves. This phenomenon continues for higher-order plants, up to the sixth-order plant. Figures 44 through 46 illustrate a broadening of the root-location 'fingerprint' with virtually no error in the response curves. In particular, the sixth-order (Figure 46) identification yields models which deviate from the true plant by a fraction of a decibel in magnitude and only about four degrees in phase. Furthermore, the error is concentrated in higher frequencies; such error distribution is easily handled by modern control system synthesis techniques.

Error distribution changes markedly, however, when the model order is increased from six to seven. Figure 47 illustrates some of the problems associated with identifying higher-order plants. The root (pole) locations shown in Figure 47(a) indicate that a significant number of the identified plants have poles outside the unit circle, erroneously indicating an unstable system. Furthermore, the response curves for this seventh-order identification show an error distribution with much of the identification error concentrated in the lower frequencies. This 'structured uncertainty' would force the controls engineer into compromising large amounts of performance in order to ensure closed-loop stability. Here, high-order identification is a losing proposition.

6.4.2 Identification Using Unweighted-Least-Squares. The next set of experiments is performed identically to those documented in the previous section with one important exception. Namely, least-squares identification is used without the benefit of weighting. We see in Figures 48 through 53 that the lack of proper weighting exacerbates the order-dependent problems associated with the identifiability of the plant under test. In particular, the second, third, and fourth-order plant identification experiments (Figures 48 – 50) yield plant models which closely approximate

132
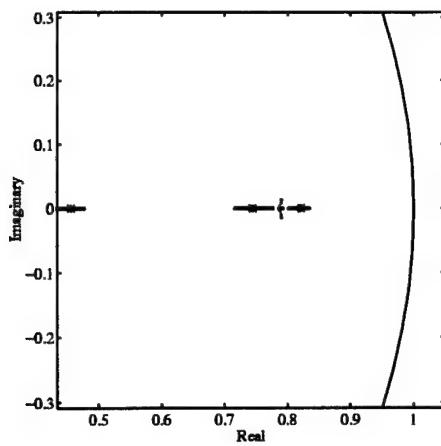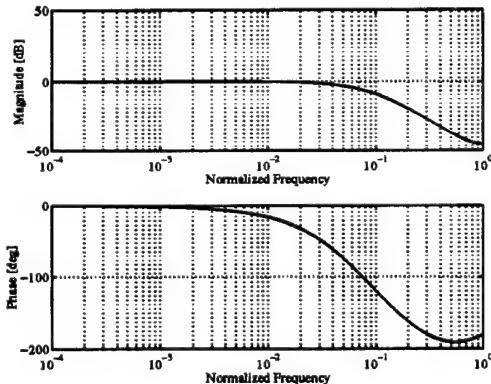
(a) Root Locations

(b) Worst-Case Envelopes
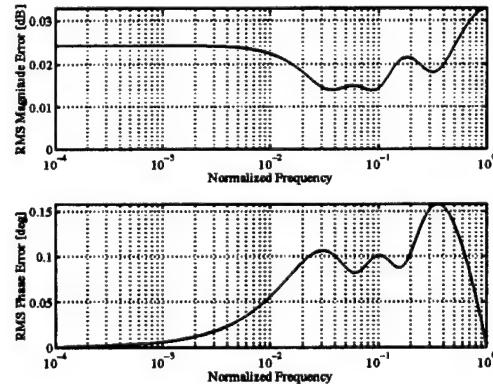
(c) RMS Errors

Figure 43. Identification Errors (Third-Order System): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

the system. In contrast, the higher-order-plant identification degrades to unacceptable error levels when the plant under test is as simple as fifth-order. Figure 51 shows a broad spread in the root-location 'fingerprint,' along with high low-frequency errors. As model order is increased to six and seven (Figures 52 and 53), the identified plants bear little resemblance to the underlying truth-model plant, again with error concentrated in the lower frequencies.

133
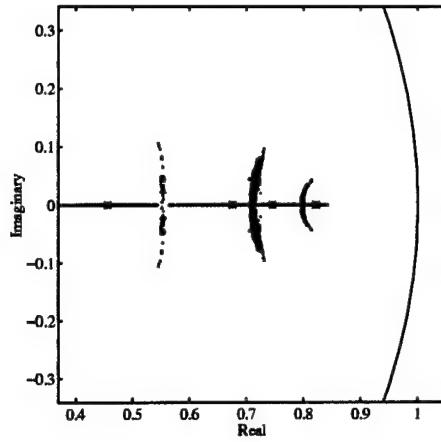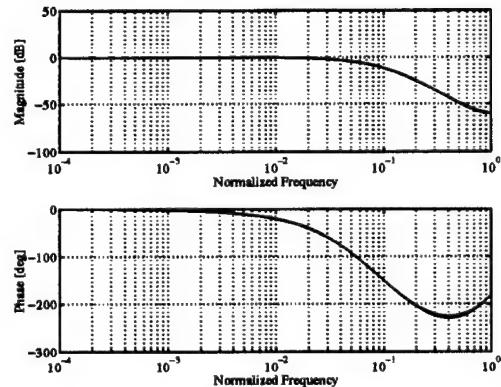
(a) Root Locations


(b) Worst-Case Envelopes
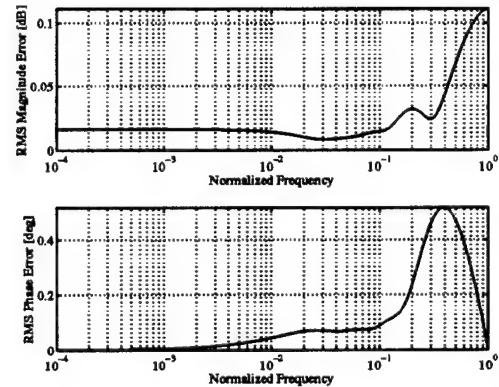
(c) RMS Errors

Figure 44.   Identification Errors (Fourth-Order System): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

## 6.5   Signal to Noise Ratio Versus ID Accuracy

The experiments documented in the previous section are all based on ID with a SNR of 100 dB. We may also wonder how varying the SNR affects the accuracy of the estimates. Surely, we expect the identification accuracy to increase as the measurements become cleaner. However, we shall see that this is not always the case.

134
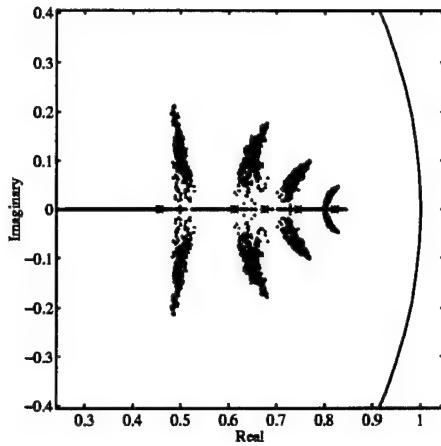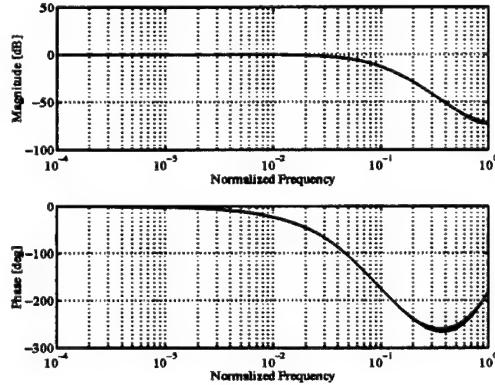
(a) Root Locations

(b) Worst-Case Envelopes
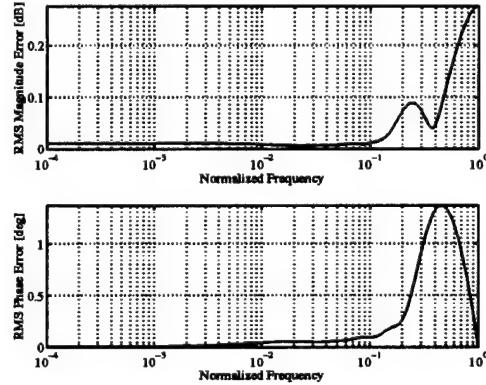
(c) RMS Errors

Figure 45.   Identification Errors (Fifth-Order System): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

Consider the ID experiment documented in Figure 47. Here, we corrupted the measurements with white noise in order to achieve a SNR of 100 dB. We now conduct the same experiment with a SNR of 40 dB. Figure 54 shows the results of this experiment.

Comparing Figures 47 and 54, we see that the identification error is lessened in the case of lower SNR! In particular, notice the reduction of low frequency error
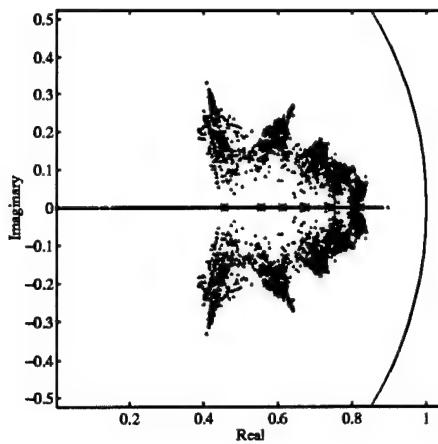
(a) Root Locations
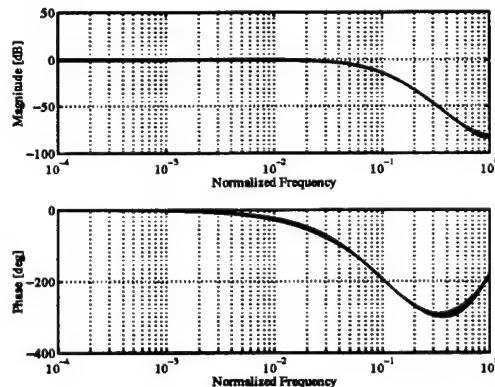
(b) Worst-Case Envelopes
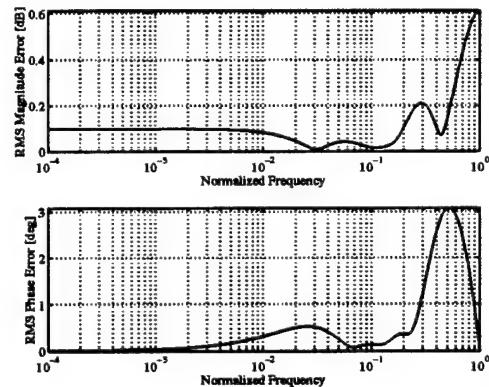
(c) RMS Errors

Figure 46. Identification Errors (Sixth-Order System): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

exhibited by the identified plants when the measurement is 40 dB below the signal. Figure 55 brings the frequency response error plots together for comparison.

Figure 55 clearly shows that the identification commits significantly greater errors when the measurements are cleaner. That is, our original hypothesis that ID errors will decrease with increasing SNR is incorrect. Now, if ID error does not
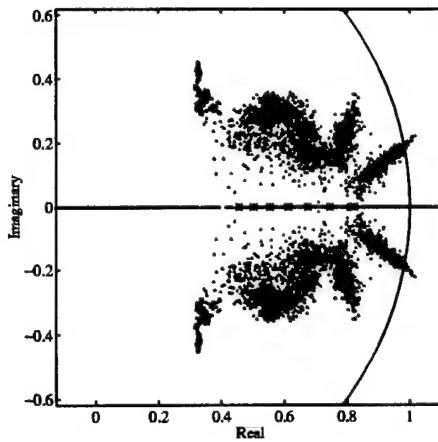
(a) Root Locations
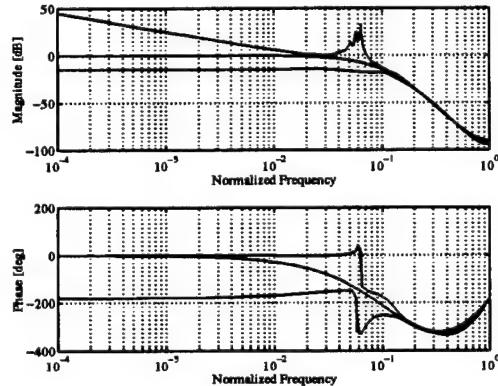

(b) Worst-Case Envelopes
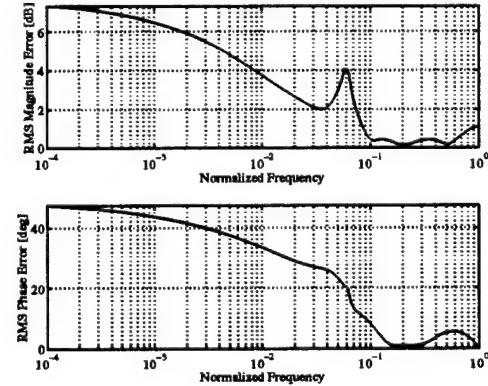
(c) RMS Errors

Figure 47.   Identification Errors (Seventh-Order System): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

monotonically decrease with growing SNR, we are motivated to determine just how the error behaves with varying SNR.

6.5.1   ID Error Versus SNR – Weighted Least Squares.   We investigate the error behavior by conducting an ID experiment with increasing SNR and observe the error committed as a function of the SNR. This experiment is carried out similarly to that for the previous case, but now we allow the SNR to range from -10 dB to 190

137
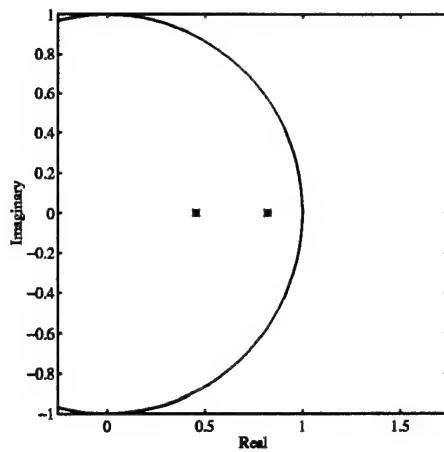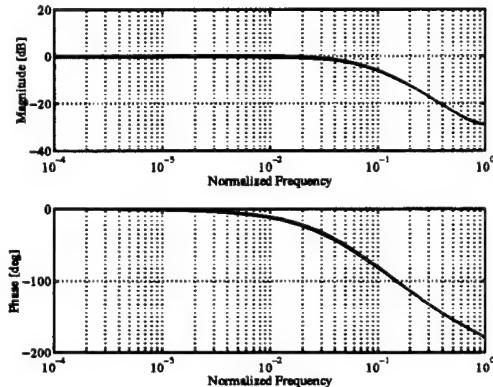
(a) Root Locations

(b) Worst-Case Envelopes
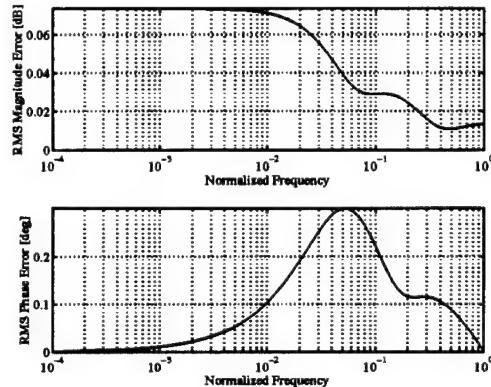
(c) RMS Errors

Figure 48.   Unweighted Least Squares Identification Errors (Second- Order System): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

dB in increments of 5 dB. At each value of SNR, we perform 1000 Weighted Least Squares identification runs in a Monte Carlo fashion (201,000 total runs). As we noted previously, the low-frequency error is most relevant to control system design, so we concentrate on the RMS error committed in the magnitude frequency response curve at three low frequencies: 10^{-4}, 10^{-3}, and 10^{-2} Normalized Frequency Units (NFU). Figure 56 shows the results of this experiment.
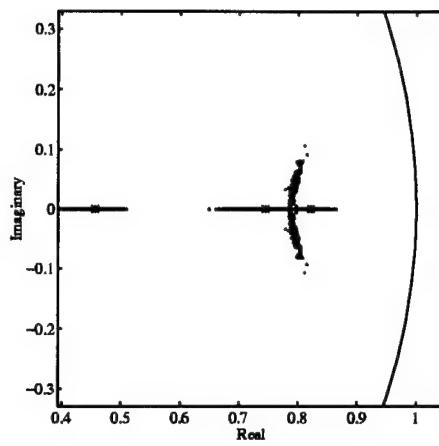
(a) Root Locations
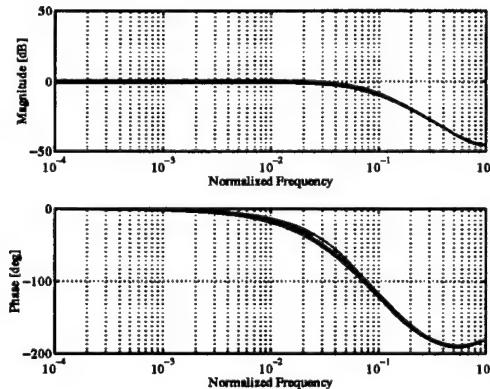
(b) Worst-Case Envelopes
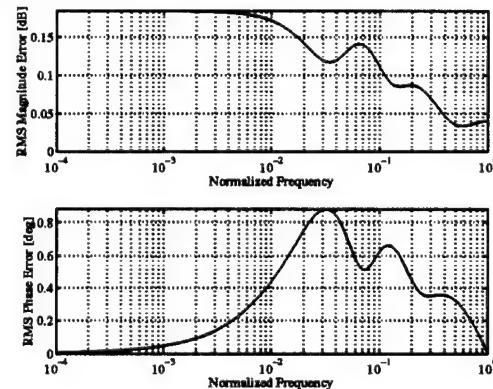
(c) RMS Errors

Figure 49. Unweighted Least Squares Identification Errors (Third-Order System): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

The plots in Figure 56 clearly show that the low frequency error exhibited by the identified models does not monotonically decrease with increasing SNR. In fact, the error decreases as expected up to about 50 dB SNR; larger SNR's after 50 dB induce a sharp rise in the low frequency response errors committed by the identified models. Thus, for this experiment, the optimum SNR is about 50 dB. Furthermore, the error effectively levels off for SNR's above about 120 dB.
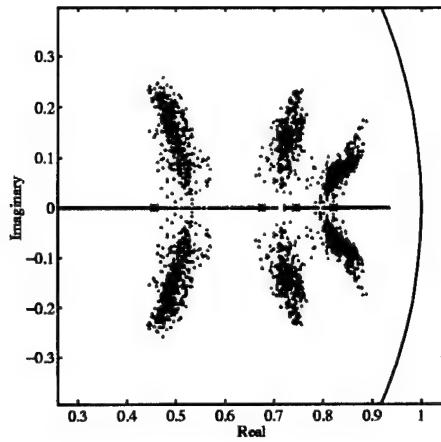
(a) Root Locations
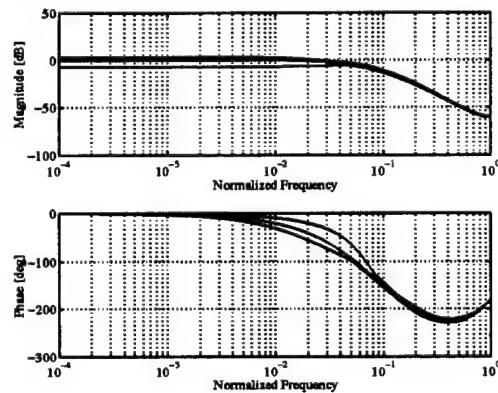

(b) Worst-Case Envelopes
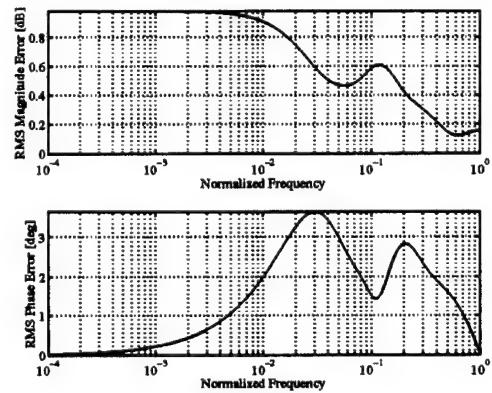
(c) RMS Errors

Figure 50.   Unweighted Least Squares Identification Errors (Fourth- Order System): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

Why do we see this strange behavior of the ID error? We propose an explanation involving the weighting used in the identification. The weighting matrix (as described in Chapter III) is derived based on the assumption of white noise on the output measurements. As SNR increases above about 50 dB, the signal begins to overpower the noise until the measurement noise is insignificant. Effectively, the ID is tuned for noise when very little noise is present. That is, the identification is mistuned, analogous to an improperly tuned Kalman filter. As SNR approaches about

140
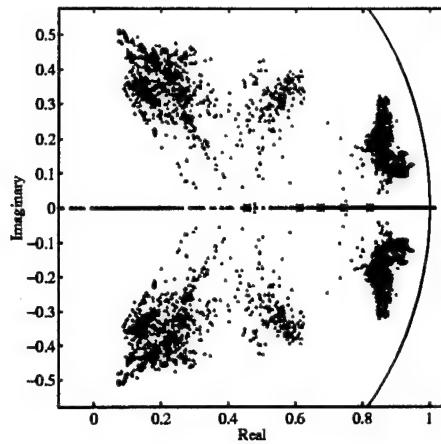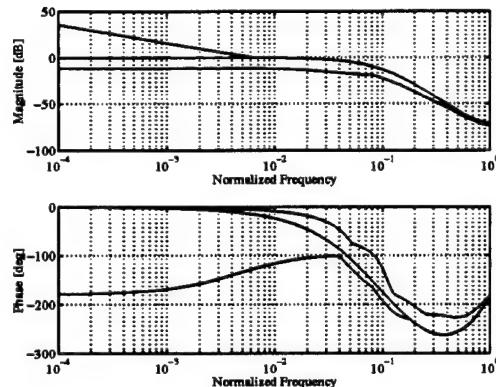
(a) Root Locations

(b) Worst-Case Envelopes
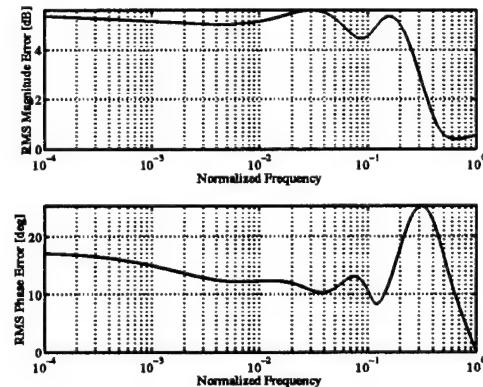
(c) RMS Errors

Figure 51. Unweighted Least Squares Identification Errors (Fifth-Order System): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

120 dB, the measurement noise is no longer a player, and numerical error drives the ID error.

6.5.2 ID Error Versus SNR – Unweighted Least Squares.   In order to verify the proposed explanation for non-decreasing ID error versus SNR, we present the results of another experiment for which the identification is accomplished with no weighting. Figure 57 shows the results of this experiment. These plots show the RMS

141
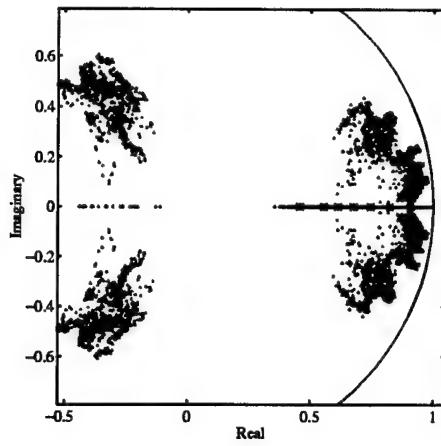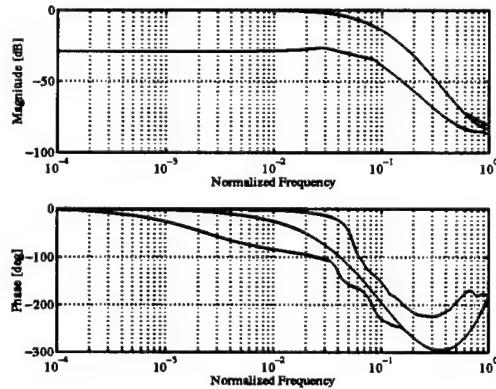
(a) Root Locations



(b) Worst-Case Envelopes
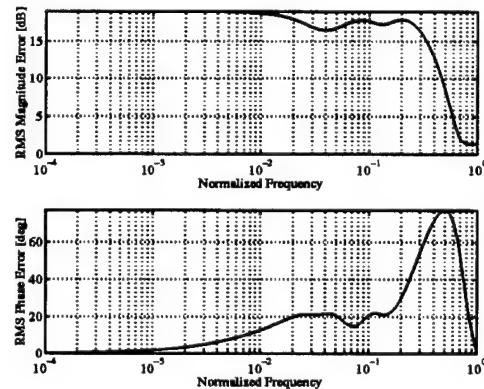
(c) RMS Errors

Figure 52.   Unweighted Least Squares Identification Errors (Sixth-Order System): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

ID error in the magnitude frequency response curves using Standard Least Squares estimation; the Weighted Least Squares results are superimposed for comparison. Here we see that the estimation error is high for low SNR and drops dramatically for SNR above about 90 dB. In contrast, the Weighted Least Squares curves show significantly better performance at low values for SNR (i.e. significantly high measurement noise). As the SNR increases to above about 140 dB, the Unweighted Least Squares estimate error drops below the error committed by the Weighted Least Squares ID.

142
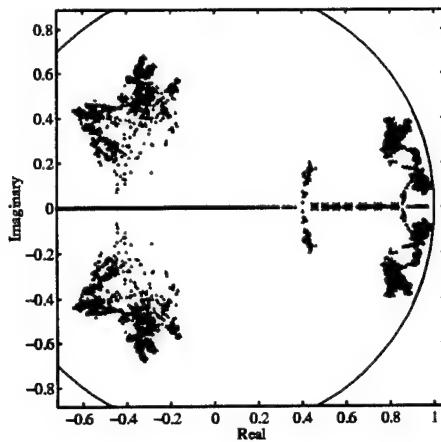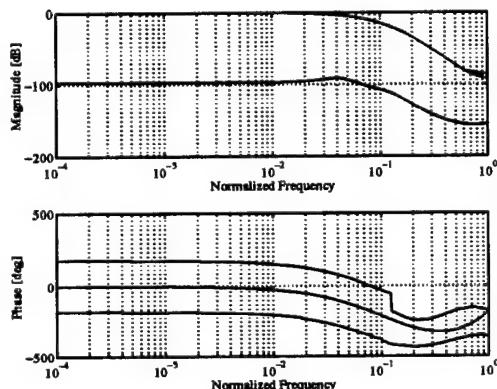
(a) Root Locations


(b) Worst-Case Envelopes
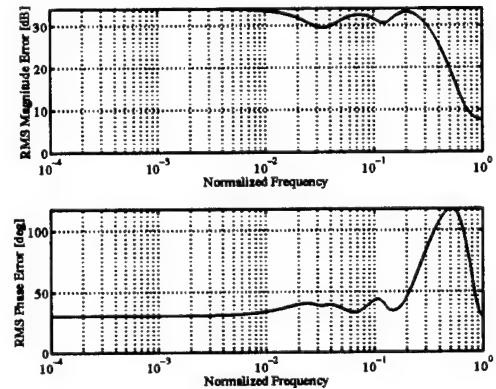
(c) RMS Errors

Figure 53.   Unweighted Least Squares Identification Errors (Seventh-Order System): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

Thus, properly tuned ID outperforms ID with incorrect measurement noise assumptions.

The phenomenon by which estimation performance drops off with increasing SNR is not without precedent. For example, a paper by Denham [15] documents the derivation of an iterated extended Kalman filter applied to a nonlinear plant with measurement noise. Denham found that the performance of the filter could be in-

143
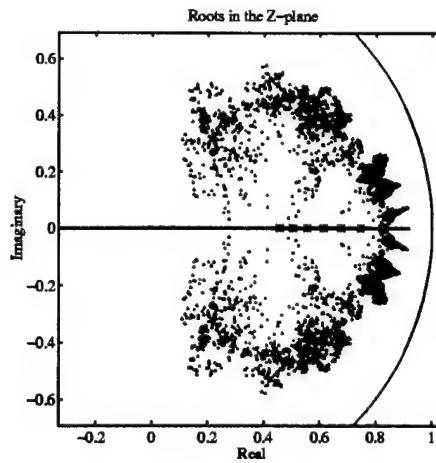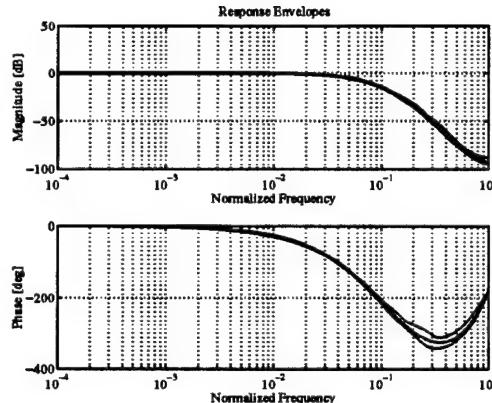
(a) Root Locations


(b) Worst-Case Envelopes
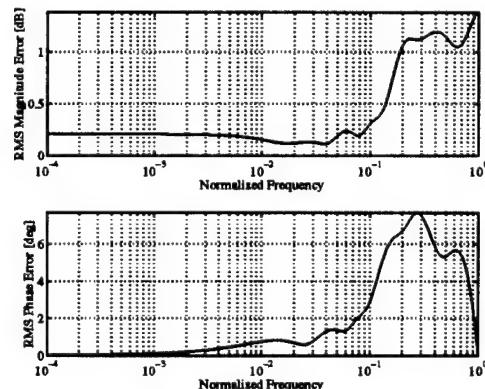
(c) RMS Errors

Figure 54.   Unweighted Least Squares Identification Errors (Seventh-Order System With 40 dB SNR): *These plots are derived from a 1000 run Monte Carlo analysis. (a) shows root locations (Actual *, Identified ·). (b) shows worst-case frequency response envelopes. (c) shows magnitude and phase RMS errors.*

creased by injecting artificial measurement noise, thereby decreasing the SNR (not to be confused with 'pseudo-noise' used in the tuning of a Kalman filter). Furthermore, the following section presents a discussion of a simple example which illustrates the potential for deleterious effects on ID performance caused by increasing SNR.

144
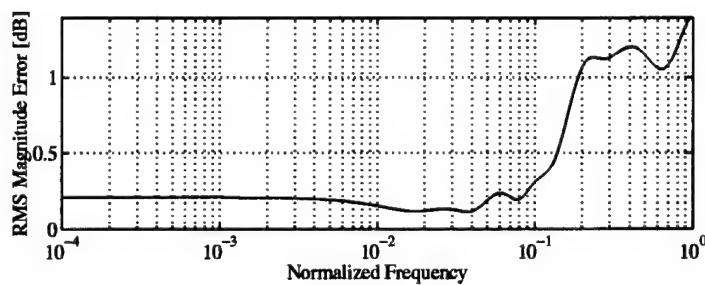
(a) SNR = 40 dB
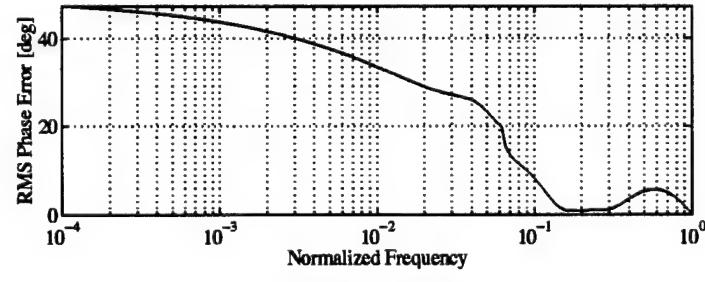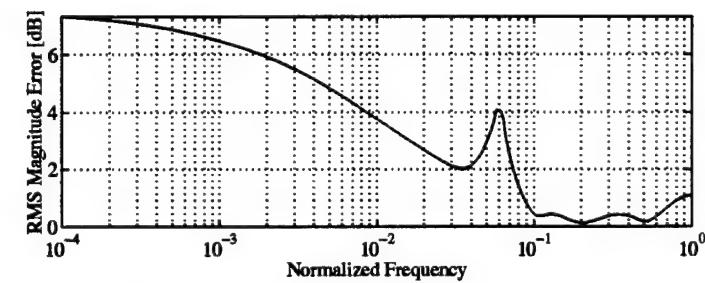
(b) SNR = 100 dB

Figure 55. Comparison of 40 dB and 100 dB SNR: *These plots show the RMS error in magnitude and phase for a seventh-order system identified via Weighted Least Squares. (a) shows the errors for a SNR of 40 dB while (b) is for 100 dB SNR.*
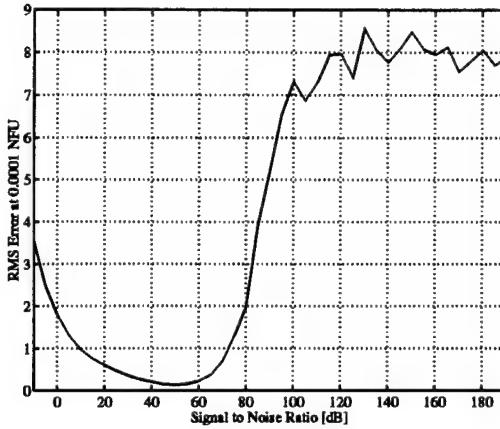
(a) RMS Error at 10^{-4} NFU
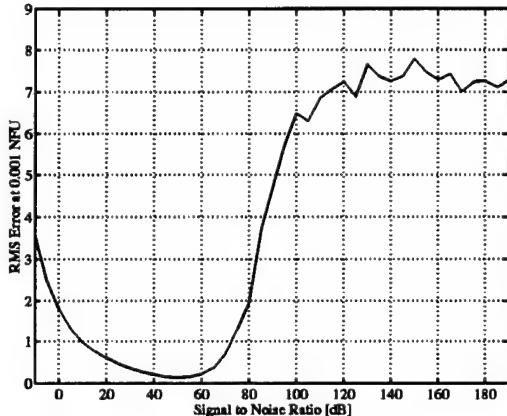
(b) RMS Error at 10^{-3} NFU
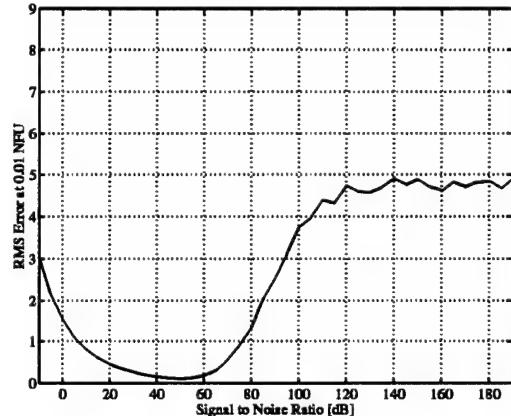
(c) RMS Error at 10^{-2} NFU

Figure 56. RMS ID Error Versus SNR: *These plots show the error committed by Weighted Least Squares as a function of SNR. The truth model is a seventh-order all real pole plant described in Section 6.4*

6.5.3 Signal to Noise Ratio Effects. We present here an interesting example which illustrates that non-judiciously increasing the SNR (e.g. by increasing the input's magnitude) may not be the best approach to enhancing ID. Thus, consider the scenario in which we wish to determine the value of an unknown static gain. With an input μ and resulting output y, the gain is given by

$$k = \frac{y}{\mu}$$

146
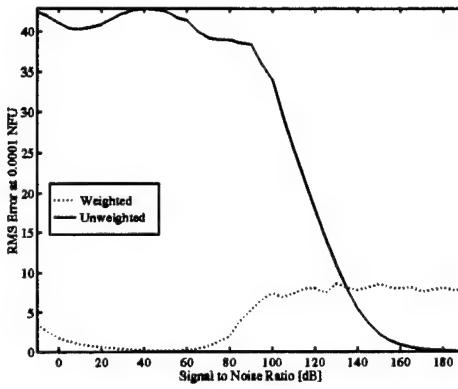
(a) RMS Error at 10^{-4} NFU
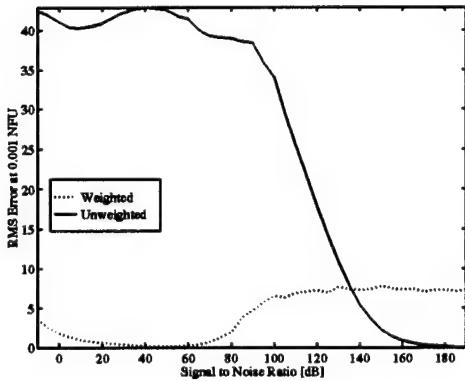
(b) RMS Error at 10^{-3} NFU
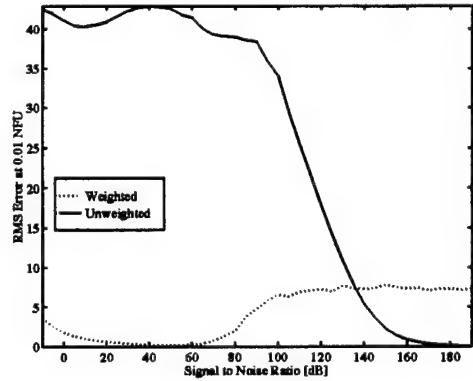
(c) RMS Error at 10^{-2} NFU

Figure 57.   RMS ID Error Versus SNR Unweighted Least Squares: *These plots show the error committed by Weighted and Unweighted Least Squares as a function of SNR. The truth model is a seventh-order all real pole plant described in Section 6.4*

We have access to noise-corrupted measurements of the output z and input u. Thus, we form the estimate of the gain by

$$\widehat{k} = \frac{z}{u} = \frac{y + v_y}{\mu + v_\mu}$$

where v_y and v_μ are realizations of the random variables v_y and v_μ.

147

Now, let us assume that v_μ and v_y are independent, zero-mean, Gaussian random variables with equal variance σ^2. With these assumptions, we calculate the expected value of the estimate:

$$\mathcal{E}\{\widehat{k}\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{y+v_y}{\mu+v_\mu}\frac{1}{2\pi\sigma^2}\exp\left\{\frac{-(v_\mu^2+v_y^2)}{2\sigma^2}\right\}dv_\mu dv_y$$

A little manipulation, aided by Mathematica® [93] allows us to evaluate the integral and yields

$$\mathcal{E}\{\widehat{k}\} = \frac{y}{\mu}\left[\frac{\sqrt{2}|\mu|}{\sigma\exp\left\{\frac{\mu^2}{2\sigma^2}\right\}}\int_0^{\frac{|\mu|}{2\sigma}}e^{t^2}dt\right]$$

Defining the SNR as $S = \frac{\mu^2}{\sigma^2}$ and recognizing that $k = \frac{y}{\mu}$, we have

$$\mathcal{E}\{\widehat{k}\} = k\left[\sqrt{2S}\exp\left\{\frac{-S}{2}\right\}\int_0^{\sqrt{\frac{S}{2}}}e^{t^2}dt\right] \tag{132}$$

Thus, we see that the expected value of the gain estimate consists of the product of the correct gain and a multiplicative bias term.

Let us concentrate on the bias term:

$$B(S) \triangleq \frac{\sqrt{2S}\int_0^{\sqrt{\frac{S}{2}}}e^{t^2}dt}{e^{\frac{S}{2}}} \tag{133}$$

$B(S)$ is nominally equal to unity – that is, if the expected value of the estimate is to be equal to the correct value of the gain, then the multiplicative bias term must reduce to one. First, we expect $B(S)$ to approach one as the SNR, S, grows very large. To confirm this, we take the limit of $B(S)$ as S approaches infinity.

$$\lim_{S\to\infty} B(S) = \lim_{S\to\infty} \frac{\sqrt{2S}\int_0^{\sqrt{\frac{S}{2}}}e^{t^2}dt}{e^{\frac{S}{2}}} \tag{134}$$

148

Repeated use of L'Hôpital's rule and Leibnitz's rule in Equation (134) yields

$$
\lim_{S \to \infty} B(S) = \lim_{S \to \infty} \frac{\frac{d}{dS}\left\{\sqrt{2S}\int_0^{\sqrt{\frac{S}{2}}} e^{t^2}\,dt\right\}}{\frac{d}{dS}\left\{e^{\frac{S}{2}}\right\}}
$$

$$
= \lim_{S \to \infty}\left\{1 + \frac{\sqrt{2}\int_0^{\sqrt{\frac{S}{2}}} e^{t^2}\,dt}{\sqrt{S}e^{\frac{S}{2}}}\right\}
$$

$$
= 1 + \sqrt{2}\lim_{S \to \infty}\frac{\frac{d}{dS}\left\{\int_0^{\sqrt{\frac{S}{2}}} e^{t^2}\,dt\right\}}{\frac{d}{dS}\left\{\sqrt{S}e^{\frac{S}{2}}\right\}}
$$

$$
= 1 + \sqrt{2}\lim_{S \to \infty}\frac{\frac{1}{2\sqrt{2S}}e^{\frac{S}{2}}}{\frac{1}{2}\sqrt{S}e^{\frac{S}{2}} + \frac{1}{2\sqrt{S}}e^{\frac{S}{2}}}
$$

$$
= 1 + \lim_{S \to \infty}\frac{1}{S+1} = 1 \tag{135}
$$

Thus, we that as the SNR approaches infinity (i.e. no measurement noise), the expected value of the estimate approaches the correct value.

Although we know that the estimate's expected value approaches the nominal value, we still may wonder about the manner in which the approach is made. Figure 58 shows the curve defined by Equation (133). The curves are created by the *quad8* function in Matlab® which employs an adaptive recursive Newton-Cotes eight-panel rule.

Surprisingly, the estimation bias is not monotonic with respect to the SNR. Figure 58(b) indicates that the multiplicative term peaks at about 1.3 for a SNR of about 4.5. Furthermore, the estimate is unbiased for a SNR of about 1.7. Thus, if our goal is to eliminate the expected bias, we would choose an input such that the SNR is fairly small!

149
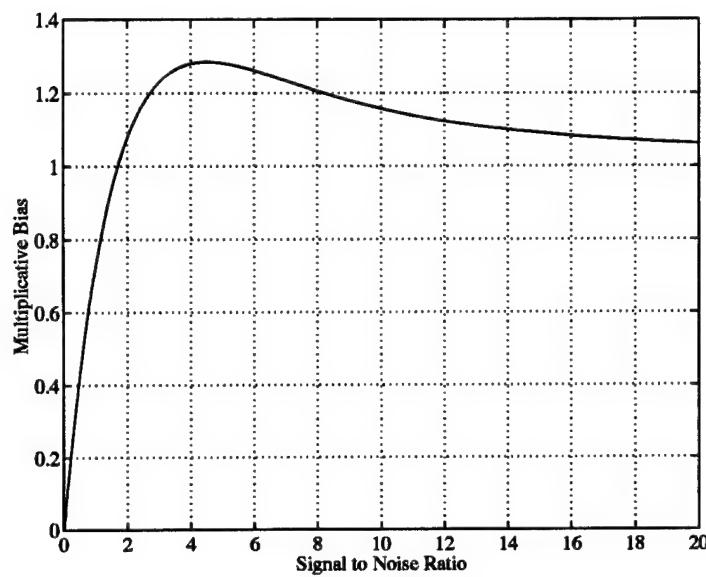
(a) Multiplicative Bias vs SNR
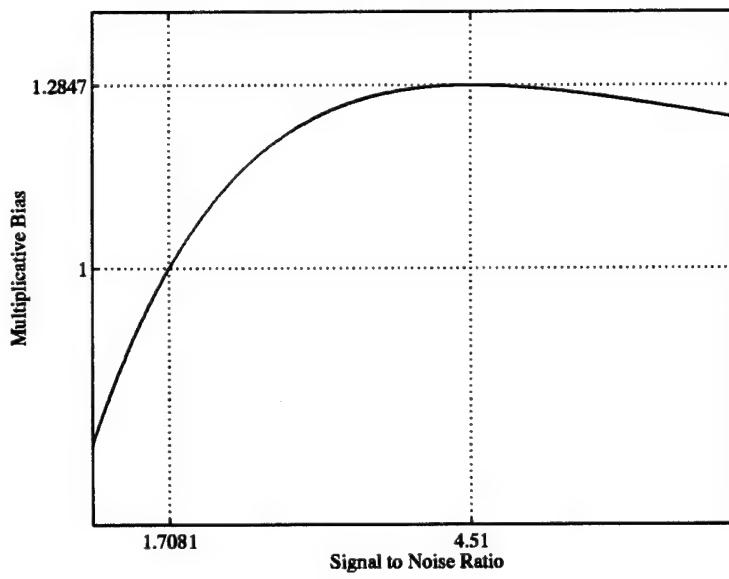
(b) Multiplicative Bias vs SNR (Detail)

Figure 58. Multiplicative Bias Versus SNR: *These plots illustrate the multiplicative term in the expected value of a gain estimate when the input and output measurements are noise-corrupted. (b) is a detail of (a).*

In conclusion, we have shown, by way of an example, that indiscriminately boosting the signal to noise ratio does not always yield the best estimates. Here, if we originally chose an input which produced a SNR of 1.7, but later decided to use more input energy, increasing the SNR to 4.5, we would actually increase the estimate bias. Obviously, we are in the realm of nonlinear estimation.

6.6 Conclusions

As we saw previously, small errors in the identified coefficients of the transfer function can produce large errors in other quantities which are important in frequency-domain control system design (e.g. locations of the plant's poles and zeros). Indeed, root position and Bode magnitude plots of the transfer function are significantly affected by seemingly small discrepancies in the coefficients.

Furthermore, over-modeling does not pay. It is better to use a simplified model of the plant rather than pursue the elusive goal of accurately capturing the high-order dynamics of the plant, and in doing so sacrifice the estimation/identification accuracy of its dominant modes.

Another interesting inference can be made from the data produced by these experiments. Namely, uncertainty of the transfer function is *not* necessarily low for low frequencies and high for higher frequencies, as is widely assumed (and required) for many robust control system design techniques [17]. The ID-induced uncertainty in the Bode plots is thus 'structured.' If robust controller design techniques are employed as a basis for the controller portion of an adaptive control system (in order to ensure robustness to high-frequency phenomena such as order reduction), the engineer must be cognizant of the structured uncertainty attributed to model coefficient errors, in addition to modeling errors resulting from model-order truncation. Since both types of uncertainty exist within the paradigm of adaptive control, the curves describing uncertainty will be complex functions of frequency. Thus, the controls engineer should remember three important facts:

1. Relatively 'small' errors in identified polynomial coefficients can produce transfer functions which deviate significantly from the nominal (in a frequency-domain context).

2. Lower-order transfer functions produce much smaller frequency-domain errors. One should not require very high-order plant models to describe the dynamics of real-world dynamical systems. Therefore, one should use the smallest order plant model possible. In fact, high-order transfer function models will almost assuredly force the ID portion of an adaptive controller to produce nonsense.

3. Errors in the coefficients of the transfer function do not necessarily manifest themselves as errors limited to high frequencies. Rather, the frequency-domain errors appear as complex functions of both the nominal transfer function and frequency. Thus, if adaptive and modern robust control design techniques are applied together, the engineer must remember to account for low-frequency errors attributed to ID error.

It has been shown that the chosen method of identification has a profound effect on the errors associated with the estimation. In this chapter, we see that a proper weighting scheme significantly enhances the ability of least-squares parameter estimation to yield acceptable plant models. Furthermore, the problems associated with identification of higher-order plants is somewhat alleviated by properly weighting the least-squares estimation.

Finally, the ID error does not necessarily decrease with increasing signal-to-noise ratio. In fact, if the identification is not properly tuned, the ID algorithm can produce errors which do not significantly decrease for arbitrarily small measurement noise. This effect is probably caused by a combination of improperly weighting the least squares estimate and numerical noise which is inherent to any computation performed on a digital computer.

ID resides in the realm of nonlinear estimation. Hence, as we have seen, non-judiciously increasing the SNR does not guarantee improvement in identification. However, the theory presented in Chapter IV provides us the mechanism for optimally distributing the input energy for any given signal strength.

So far, we have explored many aspects of input design for ID. We have developed theory and tested the theory through experiments. We move now to Chapter VII wherein we summarize this research with conclusions and recommendations for future work.

VII. Conclusions and Recommendations

7.1 Summary

In Chapter II of this dissertation, we begin with an investigation of ID in the sterilized environment facilitated by noise-free measurements. Since the required quantities were assumed to be known perfectly, we properly referred to ID in this context as modeling. The absence of measurement noise allowed us to form some conclusions concerning some requirements for proper modeling of a dynamical system. Namely, the input applied to a plant under test must meet a minimum input-order constraint. Specifically, the input must exhibit an order greater than or equal to the number of unknown parameters. Chapter II also included a potentially useful conjecture relating the determinant of the regressor matrix to the excitation order. The formula given in Conjecture 2.7.1 (on page 29) shows transparently that a necessary and sufficient condition for parameter identifiability is that the order of the input be at least equal to the length of the parameter vector. A proof of the conjecture for low-order systems is given. It is also shown that the assumed order of the model must not exceed the actual order of the true plant. A proof was presented in Chapter II which shows that over-modeling (i.e., choosing the model order too high) results in a rank-deficient regressor matrix.

Next, Chapter III dealt with the inclusion of noise added to the plant model. Here, we clearly saw that ID suffers greatly by measurement noise. The impact of measurement noise is exacerbated by the fact that this noise finds its way into the equation error in such a way that the noise terms are correlated by the very parameters the ID algorithm seeks to find. This phenomenon results in a weighting matrix (required for minimum-variance estimates) which is a function of the unknown parameters. Thus, the ID problem is nonlinear, greatly complicating the identification. Surprisingly, process noise presents less of a problem to ID. We saw that this noise is correlated only by the assumed noise model. Therefore, if we are not interested

in estimating the noise model's parameters, we can more easily account for process noise.

The problems associated with equation error correlation brought on by measurement noise were addressed via a new System Identification algorithm. Chapter III presented the derivation of this new algorithm designed around minimum variance estimation. Experiments were presented which illustrate the profound improvement offered by this algorithm which correctly accounts for measurement noise. The algorithm circumvents the paradox encountered in ID (i.e., we must know the parameters in order to estimate the parameters properly) by iteratively weighting the estimates, said weighting being derived from past estimates.

With a new ID algorithm in place, Chapter IV addressed the optimization of the inputs applied to an unknown plant for ID. This chapter presents an interesting connection between theory originally applied to communications and its application to identification. We saw that the plant, input generator, and the ID algorithm can be viewed as different entities in a communication scenario in which the message to be relayed consists of the plant parameters. This analogy allows us to formulate optimal inputs based on information theoretic concepts.

Specifically, the information contained in the input/output pairs was calculated using *entropy*. We saw in Chapter IV that entropy can be taken directly as a measure of a sequence's information content. Although an apparent dichotomy exists – entropy can be taken as a quantity to be either maximized *or* minimized for optimality – we saw that, in our context, maximizing information (entropy) is required in order to impart the maximum parameter information to the ID algorithm.

Input shaping filters were designed by maximizing output entropy while meeting several different sets of constraints. Namely, we performed the maximization while holding the average output power constant. Next, we constrained the average input power. Finally, we designed shaping filters by performing unconstrained maxi-

mization of a performance/penalty functional with negative weights applied to both input and output powers.

Each of the three approaches yielded interesting and useful results. Constraining input power while maximizing output entropy resulted in inputs with power spectral densities (PSD) which are proportional to the inverse of the plant's magnitude squared. If the measurement noise is colored, then the optimal input PSD is attenuated in frequencies for which the noise is high. In contrast, if we constrain the input power, the optimum input for ID has a PSD which is constant throughout frequency if measurement noise is ignored. Accounting for measurement noise, the optimal input PSD is attenuated within frequencies for which the noise is high and for frequencies wherein the plant exhibits low gain. Finally, the unconstrained optimization problem yields an input PSD which is a complicated combination of the previous two scenarios.

After showing the basic formulations for the optimal input PSD, Chapter IV went on to formulate the solutions when the plant is unknown, using expected values in place of the known plant values. We saw that, given certain assumptions on the probability density functions describing the unknown parameters, the solution could be formulated by replacing instances of the known plant's magnitude with the expected value of the plant's magnitude. Thus, an adaptive, iterative input generator could be based on the best current estimate of the unknown parameters.

Finally, Chapter IV discussed some implementation issues. First, each of the solutions for the optimal input PSD's contain terms which may result in nonsensical shaping filter magnitude characteristics (e.g. imaginary magnitude). We saw some ideas on dealing with cases such as those. Second, this chapter addressed the specific situation in which it is required to classify the unknown plant as being one of a predetermined finite sets of known plants and the input generator is teamed with a Multiple Model Adaptive Estimator (MMAE). We saw that this marriage might work quite well with very little on-line computational requirements.

Experiments were documented in Chapter V which confirm the optimality of the shaping filters derived previously. We saw evidence that the optimal shaping filter's magnitude response is the inverse of the plant's (when constraints are applied to the output power). Furthermore, the experiments documented here suggested a strong link between maximizing output entropy and minimizing estimation error.

Next, Chapter VI explored the limits of System Identification with respect to the utility of the identified model. Since the roots of a high-order polynomial can be very sensitive to small errors in the coefficients defining the polynomial, we saw that ID must be very accurate if we desire high-order models. In fact, the experiments documented in this chapter provide compelling impetus for limiting the model's order to the smallest usable size. Furthermore, Chapter VI explored an interesting phenomenon in which the estimation error does not necessarily monotonically decrease as signal-to-noise ratio increases. Although this behavior is initially surprising, we offered one possible explanation. Namely, the ID which is tuned for a (realistic) scenario in which measurement noise is present incorrectly weights the estimates when SNR drops to such a level as to contradict the initial assumption of measurement noise and the way it enters the process. Of course, this high level of SNR is arguably unrealistic. Further exploration of the non-monotonic behavior of estimation error with respect to SNR was carried out analytically for the simple case of estimating a gain in the presence of noisy data. This analysis showed that a relatively low SNR produces *perfect* results (in the expected value), while higher signal levels can actually result in higher expected errors.

7.2 Conclusions

First, and foremost, we must conclude that System ID accuracy is *highly* dependent on the quality of the data taken from the measured system. Although this statement seems rather obvious, engineers often neglect to account for, and properly handle, the effects of measurement noise.

The accuracy of the identification can be optimized via the proper use of weighting in the estimation and by judicious application of inputs. That is, we can significantly increase the performance of the ID algorithm by shaping the inputs applied to the system and by tailoring the ID algorithm to account for noise in the measured plant outputs and inputs.

Regardless of the input applied to the plant, ID cannot be accomplished successfully if the assumed model order is too high. We saw in Chapter II that identification breaks down if the chosen model order is greater than the order of the unknown plant. Furthermore, the probing input's order must be greater than the number of parameters to be resolved from the input/output data.

Deterministic modeling showed us the importance of proper model and input order choice. However, the choice of the input and structure of the ID profoundly affects the accuracy of the ID when measurement noise is present in the data. This noise sends the ID into the realm of stochastic estimation, thus motivating the use of a minimum-variance identification scheme which, on the surface, seems linear. However, more careful inspection shows that the weighting required to minimize the parameter error variance results in a set of equations which are nonlinear in the unknowns. Thus, the problem becomes much more difficult to solve and characterize; after all, System ID resides in the realm of nonlinear filtering. Chapter III presented and tested an iterative algorithm for computing the unknown parameters which properly accounts for measurement noise while maintaining the linear flavor of the problem, facilitating the use of well-known and easily implemented components of linear algebra. This algorithm represents a significant contribution to the field of System Identification.

Regardless of the ID algorithm used, the input applied to the unknown plant plays an important part in the accuracy of the parameter estimates. The parameter error is seen to be closely linked to the information content of the input/output plant data. This information can be manipulated by the statistics of the input, allowing

158

us to optimize (maximize) the joint information between the input/output pairs and the unknown parameter vector. Since the optimization is carried out with respect to the PSD characteristics of the output data stream, the optimal input is described by a power spectral density. Thus, the optimal input is realized by a white sequence driving a shaping filter. Although the filter is described in part by the unknown plant parameters, we can iteratively define the filter, allowing it to evolve as the parameter estimates mature throughout the life of the identification.

Experimental evidence corroborates the theory developed in this dissertation. We saw that experimentally derived suboptimal inputs' PSDs approach the PSD characteristics of the predicted optimal inputs. Furthermore, experiments indicate that proper weighting of the parameter estimates is vital to the accuracy of the identification, especially when root location is used as a metric of ID performance. Perhaps more significant, experiments indicate that parameter uncertainty can manifest itself in the frequency domain as high errors in low-frequency intervals. Thus, modern, robust, control system synthesis techniques should be applied with care as the basis of the controller-design portion of an adaptive control system.

7.3 Contributions of this Dissertation

This research contributes significantly to several different disciplines: system identification, adaptive control, and fault detection and isolation for reconfigurable control. The specific contributions are listed below:

1. Optimal (with respect to output-information content) input power spectral density is derived for any arbitrarily colored measurement noise and plant combination. Thus, we are not confined to a white measurement noise model. Furthermore, the derivations allow us to consider constraints on either the input power or the output power. Additionally, the optimal input is derived for the unconstrained case, with weighting applied to the input and output power.

159

2. An efficient ID algorithm is presented which correctly accounts for measurement noise. The algorithm adapts to changing parameter estimates in order to weight properly the equation error induced by the measurement noise and the dynamics of the system under test. Experiments show the importance of proper measurement noise accounting in the identification. No proofs of convergence are presented; rather, experimental evidence indicates the superiority of this algorithm's performance relative to the performance of a widely-used unweighted least-squares algorithm.

3. Theory and experiments are presented which underscore the nonlinearity of the estimation process. Said nonlinearity produces results which may surprise the engineer if he makes linear assumptions.

7.4 Recommendations for Future Research

The research presented in this dissertation sets the stage for several extensions. First, we considered only SISO plants for this work. The algorithm described in Appendix A should be extended to the MIMO case.

Second, the optimal input sequence derivations could be extended to the MIMO plant. Although not trivial, the research documented here should form a solid basis for this augmentation. In the MIMO case, each input sequence would not be independent. Rather, each spectrum and cross-spectrum would become germane to the derivation.

Third, the ID algorithm's performance should be evaluated within the context of model-order reduction. That is, models should be identified which are lower in order than the truth model. This extension is important because true dynamical systems are actually described by very high order transfer functions.

Perhaps the most challenging and rewarding extension to this research involves the inclusion of the effects of control inputs to the optimization. This dissertation was based on *open-loop* identification, open-loop in the sense that the only inputs

160

directly analyzed originated in the input generator. In a closed-loop scenario, inputs to the plant are a combination of disturbances, the probing signals, reference signals, and a conditioned history of the plant outputs. Since the goal here was to enhance identification, no emphasis was placed on regulating the output (beyond constraining the output power). In contrast, a closed-loop approach would include weights on ID performance, input power, output power, *and* system performance. Furthermore, the controller would see the auxiliary inputs as disturbances. Thus, the controller must be designed with these probing inputs in the budget, allowing some of this perceived disturbance to manifest in the output, thereby enhancing the identification without seriously degrading the closed-loop performance.

Appendix A. Iterative Minimum-Variance Algorithm

This appendix contains a pseudo-code rendition of the Weighted Least Squares algorithm developed for this research. Incorporated in the algorithm are two levels of recursion. First, the algorithm allows for sliding a window of data through time, initializing the weighting matrix using the parameter estimates calculated from the previous data window. Second, the parameter estimates are used iteratively to update the weighting matrix for each time window.

A.1 Notation

A.1.1 Matrices and Vectors.   All vectors are column and are denoted by lower-case boldface, with individual components denoted by the vector's designating letter in normal-face. For example, v_j is the jth component of the vector $\mathbf{v}$. Matrices are denoted by upper-case boldface (e.g. $\mathbf{H}$). Superscript T $((\cdot)^T)$ denotes matrix transposition. Superscript -1 $((\cdot)^{-1})$ denotes matrix inversion. Iteration count is given by a left-hand subscript (e.g. $_\ell\widehat{\boldsymbol{\theta}}$ denotes the ℓth iteration of $\widehat{\boldsymbol{\theta}}$). Parameter estimates are denoted by a hat $(\widehat{(\cdot)})$.

A.1.2 Sequences.   A right-hand subscript integer denotes the time index (e.g. y_k represents the sequence y evaluated at the kth time). A right-hand subscript integer applied to a vector denotes that vector evaluated at the time given by the subscript. A superscript integer in parentheses with a right-hand subscript integer applied to a vector denotes a history of a sequence arranged in a vector evaluated at

162

the time given by the subscript and having a length given by the superscript. For example,

$$
\mathbf{y}_k^{(q)} = \begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-q+1} \end{bmatrix}
$$

is the vector of length q which contains elements of y recorded at the time instants $k, k-1, \ldots, k-q+1$.

A.2 Class of System Models

The plant to identify is single-input-single-output, linear, time-invariant, finite-dimensional, and operates in discrete time, viz., its transfer function is

$$
G(z) = \frac{b_1 z^{-1} + b_2 z^{-2} + \cdots + b_m z^{-m}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}}
$$

Alternatively, the plant is described by the difference equation

$$
y_k = -a_1 y_{k-1} - \cdots - a_n y_{k-n} + b_1 u_{k-1} + \cdots + b_m u_{k-m}
$$

A.3 Identified Parameters

Referring to the description of the system model above, the System Identification algorithm generates estimates of the $n + m$ dimensional parameter vector

$$
\boldsymbol{\theta} = \begin{bmatrix} a_1 & \cdots & a_n & b_1 & \cdots & b_m \end{bmatrix}^{\mathrm{T}}
$$

A.4 Inputs

The inputs to the System Identification algorithm are sequences which represent the system input and output time histories.

A.4.1 System Inputs.   The inputs to the system are sequences of length $q + m - 1$:

$$\begin{bmatrix} u_{k-1} & u_{k-2} & \cdots & u_{k-m-q+1} \end{bmatrix}^{\mathrm{T}}$$

These inputs are known exactly by the System Identification algorithm.

A.4.2 System Outputs.   The system outputs used for System Identification are corrupted by noise. That is, the Identification algorithm processes output sequences of the past $q + n$ plant outputs corrupted by independent, additive, white noise:

$$\begin{bmatrix} z_k & z_{k-1} & \cdots & z_{k-n-q+1} \end{bmatrix}^{\mathrm{T}}$$

where

$$z_\ell = y_\ell + v_\ell, \quad \ell = k, \ k - 1, \ \ldots, \ k - n - q + 1$$

and

$$\left. \begin{aligned} \mathcal{E}\{v_\ell^2\} &= \sigma_v^2 \\ \mathcal{E}\{v_\ell v_j\} &= 0 \end{aligned} \right\} \quad \forall \ \ell, j \ni \ell \neq j$$

A.5 Prior Information

- The degree of the plant $(n - 1)$ and the order of the numerator (m) must be known.

- The intensity of the measurement noise (σ_v) is *not* required for calculation of parameter estimates. However, σ_v is needed for computation of the estimate of the parameter estimation error covariance.

- Prior knowledge of the parameter values (though not required) can be used to initialize the algorithm.

164

A.6 Algorithm

1. **Initialization**

   1.1. Choose the window size, q. Suggest $q \geq 2(n + m)^2$.

   1.2. Set iteration stopping criteria:

      1.2.1. Set $i_{\max}$ to the maximum allowable number of iterations.

      1.2.2. Set δ_θ to represent the smallest change in the parameter estimate to be recognized.

   1.3. Initialize the time index count:

   $$k \longleftarrow k_0$$

   where $k_0 \geq \max(q + m - 1, q + n)$

   1.4. Initialize the parameter estimate vector:

      1.4.1. With no prior knowledge of the parameters[1]:

      $$\hat{\boldsymbol{\theta}}_{k-1} \longleftarrow \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

      1.4.2. With prior knowledge of the parameters:

      $$\hat{\boldsymbol{\theta}}_{k-1} \longleftarrow \begin{bmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_n \\ \hat{b}_1 \\ \vdots \\ \hat{b}_m \end{bmatrix}_{\text{Initial}}$$

   1.5. Initialize vectors $\mathbf{z}_k^{(q)}$, $\mathbf{z}_{k-1}^{(q)}, \ldots, \mathbf{z}_{k-n}^{(q)}, \mathbf{u}_{k-1}^{(q)}, \mathbf{u}_{k-2}^{(q)}, \ldots, \mathbf{u}_{k-m}^{(q)}$ with the first $q + m - 1$ system inputs and $q + n$ system output measurements.

   ---

   [1]This initialization of the parameters causes the first iteration of the parameter estimate to be an unweighted least squares estimate. The weighting matrix described in step 2.4.2 becomes the identity matrix.

165

2. **Outer loop**: This loop is used to increment the data window through time.

   2.1. Reset the iteration count:

   $$i \longleftarrow 0$$

   2.2. Initialize the parameter estimate:

   $$_0\widehat{\boldsymbol{\theta}}_k \longleftarrow \widehat{\boldsymbol{\theta}}_{k-1}$$

   2.3. Build regressor matrix, $\mathbf{H}$:

   $$\mathbf{H} \longleftarrow \begin{bmatrix} -\mathbf{z}_{k-1}^{(q)} & -\mathbf{z}_{k-2}^{(q)} & \cdots & -\mathbf{z}_{k-n}^{(q)} & \mathbf{u}_{k-1}^{(q)} & \mathbf{u}_{k-2}^{(q)} & \cdots & \mathbf{u}_{k-m}^{(q)} \end{bmatrix}$$

   2.4. **Inner loop**: This loop is used to iterate on the data for a fixed window (at time k).

      2.4.1. Increment the iteration counter:

      $$i \longleftarrow i + 1$$

      2.4.2. Build the weighting matrix $(_i\mathbf{W})$:

      2.4.2.1.

      $$\begin{cases} \mu_0 & \longleftarrow & 1 + \sum_{\ell=1}^{n}(_i\widehat{a}_\ell)^2 \\ \mu_1 & \longleftarrow & _i\widehat{a}_1 + \sum_{\ell=1}^{n-1}(_i\widehat{a}_\ell)(_i\widehat{a}_{\ell+1}) \\ & \vdots & \\ \mu_j & \longleftarrow & _i\widehat{a}_j + \sum_{\ell=1}^{n-j}(_i\widehat{a}_\ell)(_i\widehat{a}_{\ell+j}) \\ & \vdots & \\ \mu_n & \longleftarrow & _i\widehat{a}_n \end{cases}$$

      2.4.2.2.

      $$\begin{cases} \mathbf{R} \longleftarrow \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_n & 0 & \cdots & 0 \\ \mu_1 & \mu_0 & \mu_1 & \cdots & \mu_n & \ddots & \vdots \\ \vdots & \mu_1 & \mu_0 & \mu_1 & & \ddots & 0 \\ \mu_n & \vdots & \mu_1 & \ddots & \ddots & & \mu_n \\ 0 & \mu_n & & \ddots & \mu_0 & \mu_1 & \vdots \\ \vdots & \ddots & \ddots & & \mu_1 & \mu_0 & \mu_1 \\ 0 & \cdots & 0 & \mu_n & \cdots & \mu_1 & \mu_0 \end{bmatrix} \end{cases}$$

2.4.2.3. $_i\mathbf{W} \longleftarrow \mathbf{R}^{-1}$ (matrix inverse)

2.4.3. Calculate the parameter estimate error covariance matrix[2]:

$$_i\widehat{\mathbf{P}} \longleftarrow \left[\mathbf{H}^{\mathrm{T}}(_i\mathbf{W})\mathbf{H}\right]^{-1}$$

2.4.4. Generate the parameter estimates:

$$_i\widehat{\boldsymbol{\theta}}_k \longleftarrow \widehat{\mathbf{P}}\mathbf{H}^{\mathrm{T}}(_i\mathbf{W})\mathbf{z}_k^{(q)}$$

2.4.5. Check need for more iterations.

2.4.5.1. If $||_i\widehat{\boldsymbol{\theta}}_k - {}_{i-1}\widehat{\boldsymbol{\theta}}_k|| > \delta_\theta$ and $i < i_{\max}$ then go to Step 2.4.

2.4.5.2. Else finish iterating:

2.4.5.2.1. Store current estimate:

$$\widehat{\boldsymbol{\theta}}_k \longleftarrow {}_i\widehat{\boldsymbol{\theta}}_k$$

2.4.5.2.2. Continue with Step 2.5

2.5. Check for more data.

2.5.1. If more data exist

2.5.1.1. Increment k: $k \longleftarrow k + 1$.

2.5.1.2. Shift data to make room for next measurements:

$$z_{k-j} \quad \longleftarrow \quad z_{k-j+1}, \quad j = n + q - 1 \quad \text{down to} \quad 1$$
$$u_{k-j} \quad \longleftarrow \quad u_{k-j+1}, \quad j = m + q - 1 \quad \text{down to} \quad 1$$

2.5.1.3. Load z_k.

2.5.1.4. Load u_k.

2.5.1.5. Goto step 2.

2.5.2. Else terminate.

---

[2]The parameter estimate error covariance matrix estimate is actually given by $\sigma_v^2\widehat{\mathbf{P}}$. Since this represents modification by a scalar multiple (σ_v^2), the scalar cancels in the calculation of the parameter estimates; thus σ_v is *only* required for an accurate estimate of the estimation error covariance matrix.

Appendix B. Existence of $\mathcal{E}_{\boldsymbol{\theta}}\{|\mathbf{G}(\omega)|^2\}$

The plant's transfer function, $\mathbf{G}$, used in this research is described by a set of parameters, denoted $\boldsymbol{\theta}$:

$$
\begin{aligned}
\mathbf{G}(z) &= \frac{b_1 z^{-1} + b_2 z^{-2} + \cdots + b_n z^{-n}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}} \\
&= \frac{b_1 z^{n-1} + b_2 z^{n-2} + \cdots + b_n}{z^n + a_1 z^{n-1} + a_2 z^{n-2} + \cdots + a_n}
\end{aligned}
\tag{136}
$$

where

$$
\boldsymbol{\theta} \triangleq \begin{bmatrix} a_1 & \cdots & a_n & b_1 & \cdots & b_n \end{bmatrix}
$$

Now, the frequency response of the transfer function given in Equation (136) is found by evaluating $\mathbf{G}(z)$ at $z = e^{j\omega\pi}$ and evaluating the magnitude of the resulting complex number, where ω is taken to be in Normalized Frequency Units (NFU) $(-1 < \omega < 1)$.

Taking uncertainty into account, we treat the $\boldsymbol{\theta}$ vector as a random vector. Thus, the transfer function becomes a random variable:

$$
\mathbf{G}(z) = \frac{\mathbf{b}_1 z^{n-1} + \mathbf{b}_2 z^{n-2} + \cdots + \mathbf{b}_n}{z^n + \mathbf{a}_1 z^{n-1} + \mathbf{a}_2 z^{n-2} + \cdots + \mathbf{a}_n}
\tag{137}
$$

We can (formally) compute the expected value of the squared magnitude of this random transfer function:

$$
\mathcal{E}_{\boldsymbol{\theta}}\{|\mathbf{G}(\omega)|^2\} = \int_{\mathbb{R}^{2n}} p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \left[\mathbf{G}(\boldsymbol{\theta}, z)\mathbf{G}(\boldsymbol{\theta}, -z)\right]_{z=e^{j\omega\pi}} d\boldsymbol{\theta}
\tag{138}
$$

where $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is the multidimensional probability density function associated with the random vector $\boldsymbol{\theta}$.

The goal of this Appendix is to show that the integral given in Equation (138) does not in general exist. Thus, we concentrate on a simple representative example.

Namely, consider the case in which $\boldsymbol{\theta}$ consists of a single element, $\mathbf{a}$, and $\mathbf{G}$ is a simple first-order transfer function:

$$\mathbf{G}(z) = \frac{K}{z + \mathbf{a}} \tag{139}$$

where K is a known constant and $\mathbf{a}$ is a random variable.

Substituting Equation (139) into Equation (138), the expected value of the squared magnitude of this transfer function is formally computed by

$$\mathcal{E}_{\boldsymbol{\theta}}\left\{|\mathbf{G}(\omega)|^2\right\} = \int_{-\infty}^{\infty} p_{\mathbf{a}}(a) \frac{K^2}{1 + a^2 + 2a\cos(\omega\pi)} da \tag{140}$$

In particular, consider the expected value for $\omega = 0$ (i.e. the expected value of the transfer function at DC). Now, (140) becomes

$$\mathcal{E}_{\boldsymbol{\theta}}\left\{|\mathbf{G}(\omega = 0)|^2\right\} = \int_{-\infty}^{\infty} p_{\mathbf{a}}(a) \frac{K^2}{(1 + a)^2} da \tag{141}$$

The integrand in Equation (141) exhibits a second-order singularity at $a = -1$ (depending on the form of $p_{\mathbf{a}}$). Thus, for many probability density functions, the integral does not exist. For example, if we assume the density function to be Gaussian, the expected value cannot be calculated. On the other hand, we could form $p_{\mathbf{a}}$ such that the integral exists. One such form would be to assume that $\mathbf{a}$ is uniformly distributed along $[0.25, 0.75]$.

The simple example given here clearly illustrates that the expected value (taken over the random parameters) of the square of the magnitude of a random transfer function does not in general exist for all frequencies. Furthermore, if the transfer function is defined by many random coefficients, the interaction between the coefficients will create singularities at many points in frequency.

Bibliography

1. Anderson, Brian D. O. "Adaptive Systems, Lack of Persistency of Excitation and Bursting Phenomena," *Automatica*, *21*(3):247–258 (1985).

2. Aoki, M. and M. Staley. "On Input Signal Synthesis in Parameter Identification," *Automatica*, *6*:431–440 (1970).

3. Arimoto, S. and H. Kimura. "Optimum Input Test Signals for System Identification – An Information-Theoretical Approach," *International Journal of Systems Science*, *1*(3):279–290 (1971).

4. Åström, Karl. J. and Pieter. Eykhoff. "System Identification – A Survey," *Automatica*, *7*:123–161 (1971).

5. Åström, Karl J. and Björn Wittenmark. *Computer-Controlled Systems*. Englewood Cliffs, New Jersey: Prentice Hall, 1990. ISBN 0-13-168600-3.

6. Barkesseh, Mohammad N., et al. "An Identification-Based Matched Filtering Method for Fault Detection." *Proceedings of the American Control Conference*. 1907–1908. May 1990.

7. Basseville, Michèle. "Detecting Changes in Signals and Systems – A Survey." *2nd IFAC Workshop on Adaptive Systems in Control and Signal Processing*. July 1986.

8. Benveniste, Albert, et al. "The Asymptotic Local Approach to Change Detection and Model Validation," *IEEE Transactions on Automatic Control, AC-32*(7):583–591 (July 1987).

9. Beyer, William H., editor. *CRC Handbook of Mathematical Sciences* (4th Edition). Boca Raton, FL: CRC Press, 1987.

10. Blahut, Richard E. *Principles and Practice of Information Theory*. Reading, Massachusetts: Addison-Wesley Publishing Company, 1987. ISBN 0-201-10709-0.

11. Caglayan, Alper K. "Necessary and Sufficient Conditions for Detectability of Jumps in Linear Systems," *IEEE Transactions on Automatic Control, AC-25*(4):833–834 (August 1980).

12. Chow, Edward Y. and Alan S. Willsky. "Analytical Redundancy and the Design of Robust Failure Detection Systems," *IEEE Transactions on Automatic Control, AC-29*(7):603–614 (July 1984).

13. Cole, Charles. "Shannon Revisited: Information in Terms of Uncertainty," *Journal of the American Society for Information Science*, *44*(4):204–211 (1993).

14. D'Azzo, John J. and Constantine H. Houpis. *Linear Control System Analysis and Design* (Third Edition). New York: McGraw-Hill, 1988.

15. Denham, Walter F. and Samuel Pines. "Sequential Estimation When Measurement Function Nonlinearity is Comparable to Measurement Error," *AIAA Journal, 4*(6):1071–1076 (June 1966).

16. Ding, X. and P. M. Frank. "Fault Detection via Optimally Robust Detection Filters." *Proceedings of the 28th Conference on Decision and Control*. 1767–1772. December 1989.

17. Doyle, John C., et al. *Feedback Control Theory*. Macmillan, 1992.

18. Elnaggar, Ashraf, et al. "System Identification and Adaptive Control Based on a Variable Regression for Systems Having Unknown Delay." *Proceedings of the 29th Conference on Decision and Control*. 1445–1450. December 1990.

19. Eykhoff, Pieter. *System Identification – Parameter and State Estimation*. John Wiley & Sons, 1974.

20. Fel'dbaum, A. A. *Optimal Control Systems, 22*. Mathmetics in Science and Engineering. New York: Academic Press, 1965. Library of Congress Catalog Card Number 65-26397.

21. Fiedlander, B. and B. Porat. "The Modified Yule-Walker Method of ARMA Spectral Estimation," *IEEE Transactions on Aerospace Electronic Systems, AES-20*(2):158–173 (March 1984).

22. Frank, P. M. "Advanced Fault Diagnosis Techniques in Aerospace Systems." *COMPEURO '89 – 3rd Annual European Computer Conference*. 3.136–3.143. May 1989.

23. Frank, Paul M. "Fault Diagnosis in Dynamic Systems Using Analytical and Knowledge-Based Redundancy – A Survey and Some New Results." *Proceedings of the IFAC/IMAC/IFORS International Symposium on Advanced Information Processing in Automatic Control*. July 1989.

24. Gawthrop, P. J., et al. "Identification of Partially-Known Systems," *Automatica, 28*(4):831–836 (1992).

25. Gevers, Michel and Lennart Ljung. "Optimal Experiment Designs with Respect to the Intended Model Applicaitons," *Automatica, 22*(5):543–554 (1986).

26. Giri, F., et al. "Parameter Estimation Aspects in Adaptive Control," *Automatica, 27*(2):399–402 (1991).

27. Golub, Gene H. and Charles F. Van Loan. *Matrix Computations*. Baltimore, Maryland: The Johns Hopkins University Press, 1983. ISBN 0-8018-3010-9.

28. Goodwin, G. C. and R. L. Payne. *Dynamic Systems Identification: Experiment Design and Data Analysis*. Academic Press, 1977.

29. Gustavsson, I., et al. "Identification of Processes in Closed Loop – Identifiability and Accuracy Aspects," *Automatica, 13*:59–75 (1977).

30. Hannan, E. J. and M. Deistler. *The Statistical Theroy of Linear Systems*. New York: John Wiley and Sons, 1988.

31. Helmicki, A. J., et al. "Least Squares Methods for H_∞ Control Oriented System Identification," *IEEE Transactions on Automatic Control*, *38*(5):819–876 (1993).

32. Herrera-Bendezú, Luis G. and James T. Cain. "A New Criterion for Optimal Parameter Identification." *Proceedings of the American Control Conference*. 1210–1215. June 1991.

33. Horak, Dan T. and Bryan H. Allison. "Isolation of Parametric Failures in Dynamic Systems." *Proceedings of the American Control Conference*. 868–873. June 1991.

34. II, James M. Brown. *Design of an Air-to-Air Missile Autopilot Using Linear Quadratic Gaussian/Loop Transfer Recovery*. MS thesis, Air Force Institute of Technology, 1990.

35. II, James M. Brown, et al. "Autopilot Design for a Tail-Controlled Missile Using LQG/LTR With Eigenstructure Reassignment." *Proceedings of the American Control Conference*. 1994.

36. Jalfon, Agnès Cohen and Yoram Halevi. "On Partially Augmented Observers for Systems with Colored Noises." *Proceedings of the 29th Conference on Decision and Control*. 210–215. December 1990.

37. Kelley, Walter G. and Allan C. Peterson. *Difference Equations: An Introduction with Applications*. San Diego, CA: Academic Press, 1991. ISBN 0-12-403325-3.

38. Kirk, Donald E. *Optimal Control Theory*. Englewood Cliffs, New Jersey: Prentice Hall, 1970.

39. Kowalczuk, Zdzislaw. "Competitive Identification for Self-Tuning Control: Robust Estimation Design and Simulation Experiments," *Automatica*, *28*(1):193–201 (1992).

40. Krause, James M., et al. "Robust Adaptive Control: Stability and Asymptotic Performance," *IEEE Transactions on Automatic Control*, *37*(3):316–331 (March 1992).

41. Larimore, Wallace E. and John Baillieul. "Identification and Filtering of Nonlinear Systems Using Canonical Variate Analysis." *Proceedings of the 29th Conference on Decision and Control*. 635–639. December 1990.

42. Ljung, Lennart. "On the Estimation of Transfer Functions," *Automatica*, *21*(6):677–696 (1985).

43. Ljung, Lennart, et al. "Identification of Linear Multivariable Systems Operating Under Linear Feedback Control," *IEEE Transactions on Automatic Control*, *AC-19*(6):836–840 (December 1974).

44. Ljung, Lennart and Torsten Söderström. *Theory and Practice of Recursive Identification*. Cambrridge, Massachusetts: The MIT Press, 1983.

45. Lopez-Toledo, Alejandro Antonio. *Optimal Inputs for Identification of Stochastic Systems*. PhD dissertation, Massachusetts Institute of Technology, August 1974.

46. Luenberger, David G. *Optimization by Vector Space Methods*. Series in Decision and Control, New York: John Wiley and Sons, Inc., 1969.

47. Marquez, H. J. and C. P. Diduch. "Sensitivity-Robustness in Failure Detection: A Frequency Domain Approach." *Proceedings of the 29th Conference on Decision and Control*. 1389–1390. December 1990.

48. Marquez, H. J. and C. P. Diduchi. "Sensitivity of Failure Detection Using Generalized Observers," *Automatica*, *28*(4):837–840 (1992).

49. Massoumnia, Mohammad-Ali. "A Geometric Approach to the Synthesis of Failure Detection Filters," *IEEE Transactions on Automatic Control*, *AC-31*(9):839–846 (September 1986).

50. Massoumnia, Mohammad-Ali, et al. "Failure Detection and Identification," *IEEE Transactions on Automatic Control*, *AC-34*(3):316–321 (March 1989).

51. Maybeck, Peter S. *Stochastic Models, Estimation and Control, I*. 111 Fifth Aveneue, New York, NY: Academic Press, 1979. ISBN 0-12-480701-1 (V.1).

52. Maybeck, Peter S. *Stochastic Models, Estimation and Control, III*. 111 Fifth Aveneue, New York, NY: Academic Press, 1982. ISBN 0-12-480703-8 (V.3).

53. Maybeck, Peter S. *Stochastic Models, Estimation and Control, II*. 111 Fifth Aveneue, New York, NY: Academic Press, 1982. ISBN 0-12-480702-X (V.2).

54. Mayne, D. Q., et al. "A New Algorithm for Recursive Estimation of Parameters in Controlled ARMA Processes," *Automatica*, *20*(6):751–760 (1984).

55. Mehra, Raman K. "Optimal Input Signals for Parameter Estimation in Dynamic Systems – Survey and New Results," *IEEE Transactions on Automatic Control*, *AC-19*(6):753–768 (December 1974).

56. Mehra, Raman K. "Optimal Inputs for Linear System Identification," *IEEE Transactions on Automatic Control*, *AC-19*(3):192–200 (June 1974).

57. Melsa, James L. and David L. Cohn. *Decision and Estimation Theory*. New York: McGraw-Hill, 1978.

58. Meyn, Sean P. and Lyndon J. Brown. "Adaptive Control Using a Kalman Filter Based Estimation Algorithm." *Proceedings of the 29th Conference on Decision and Control*. 1432–1437. December 1990.

59. Min, Paul Seungkyu. *Detection of Incipient Failures in Dynamic Systems*. PhD dissertation, The University of Michigan, 1987.

60. Ng, T. S., et al. "Optimal Input Design for an AR Model with Output Constraints," *Automatica*, *20*(3):359–362 (1984).

61. Ng, Tung Sang, et al. "Optimal Experiment Design for Linear Systems with Input-Output Constraints," *Automatica*, *13*:571–577 (1977).

62. Ng, Tung Sang and Zahid H. Qureshi. "Optimal Experiment Design for Autoregressive Model with Output Power Constraints," *IEEE Transactions on Automatic Control*, *AC-26*(3):739–745 (June 1981).

63. Norton, J. P. *An introduction to identification*. San Diego, CA 92101: Academic Press, 1986. ISBN 0-12-521730-7.

64. Norton, J. P. "Identification and Application of Bounded-Parameter Models," *Automatica*, *23*(4):497–507 (1987).

65. Norton, J. P. "Identification and Application of Bounded-Parameter Models," *Automatica*, *23*(4):497–507 (1987).

66. Olin, Peter M. and Giorgio Rizzoni. "Design of Robust Fault Detection Filters." *Proceedings of the 29th Conference on Decision and Control*. 1522–1527. December 1990.

67. Olmstead, David N. *Optimal Feedback Controls for Parameter Identification*. PhD dissertation, Air Force Institute of Technology, February 1979. Technical Report AFAL-TR-79-1022.

68. Papoulis, Athanasios. *Probability, Random Variables, and Stochastic Processes (Third Edition)*. New York: McGraw Hill, 1991. ISBN 0-07-048477-5.

69. Partington, Jonathan R. "Worst-Case Identification in ℓ_2: Linear and Nonlinear Algorithms," *Systems and Control Letters*, *22*:93–98 (1994).

70. Patton, Ron, et al. *Fault Diagnosis in Dynamic Systems – Theory and Application*. 66 Wood Lane End, Hemel Hempstead, Hertfordshire, HP2 4RG: Prentice Hall International, 1989. ISBN 0-13-308263-6.

71. Perriot-Mathonna, Dominique M. "Improvements in the Application of Stochastic Estimation Algorithms – Parameter Jump Detection," *IEEE Transactions on Automatic Control*, *AC-29*(11):962–969 (November 1984).

72. PRO-MATLAB for Sun Workstations, The MathWorks, Inc., 21 Eliot Street, South Natick, MA 01760.

73. Ralston, Anthony and Philip Rabinowitz. *A First Course in Numerical Analysis*. New York: McGraw-Hill, 1978. ISBN 0-07-051158-6.

74. Ridgely, D. Brett and Siva S. Banda. *Introduction to robust multivariable control*. Technical Report AFWAL-TR-85-3102, Wright Patterson Air Force Base, OH: Control Dynamics Branch, Flight Dynamics Laboratory, February 1986.

75. Riggins, Jr., Robert Nelson. *Detection and Isolation of Plant Failures in Dynamic Systems*. PhD dissertation, The University of Michigan, 1991.

76. Salgado, Mario E., et al. "Approximate Identification of Linear Stochastic Systems." *Proceedings of the 29th Conference on Decision and Control.* 3148–3153. December 1990.

77. Sating, Richard R. *Development of an analog MIMO Quantitive Feedback Theory (QFT) CAD Package.* MS thesis, Air Force Institute of Technology (AU), June 1992.

78. Sating, Richard R., et al. "Development of a MIMO QFT CAD Package (Version 2)." *Proceedings of the American Control Conference.* 3081–3083. 1993.

79. Schweppe, Fred C. *Uncertain Dynamic Systems.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1973. ISBN 0-13-935593-6.

80. Shannon, Claude E. and Warren Weaver. *The Mathematical Theory of Communication.* Urbana, Il: The University of Illinois Press, 1964.

81. Sheldon, Stuart N. and Peter S. Maybeck. "An Optimizing Design Strategy for Multiple Model Adaptive Estimation and Control." *Proceedings of the 29th Conference on Decision and Control.* 3522–3527. December 1990.

82. Srichander, R. and Bruce K. Walker. "Stochastic Stability Analysis for Continuous Time Fault Tolerant Control Systems." *Proceedings of the American Control Conference.* 493–501. June 1991.

83. Stanley, William D., et al. *Digital Signal Processing.* Reston Publishing Company, Inc., 1984.

84. Tugnait, Jitendra K. "Stochastic System Identification with Noisy Input Using Cumulant Statistics." *Proceedings of the 29th Conference on Decision and Control.* 1080–1085. December 1990.

85. Tulleken, Herbert J. A. F. "Generalized Binary Noise Test-Signal Concept for Improved Identification-Experiment Design," *Automatica, 26*(1):37–49 (1990).

86. Upadhyaya, B. R. and H. W. Sorenson. "Synthesis of Linear Stochastic Signals in Identification Problems," *Automatica, 13*:615–622 (1977).

87. Vajda, S., et al. "Direct and Indirect Least Squares Methods in Continuous-Time Parameter Estimation," *Automatica, 23*(6):707–718 (1987).

88. Wellstead, P. E. "Reference Signals for Closed-Loop Identification," *International Journal of Control, 26*(6):945–962 (1977).

89. White, John E. and Jason L. Speyer. "Detection Filter Design: Spectral Theory and Algorithms," *IEEE Transactions on Automatic Control, AC-32*(7):593–603 (July 1987).

90. Williamson, Geoffrey A., et al. "Locally Robust Identification of Linear Systems Containing Unknown Gain Elements with Application to Adapted IIR Lattice Models," *Automatica, 27*(5):783–798 (1991).

91. Willsky, Alan S. "A Survey of Design Methods for Failure Detection in Dynamic Systems," *Automatica*, *12*:601–611 (1976).

92. Witsenhausen, Hans S. "Separation of Estimation and Control for Discrete Time Systems," *Proceedings of the IEEE*, *59*(11):1557–1566 (November 1971).

93. Wolfram, Stephen. *Mathematica: A System for Doing Mathematics by Computer* (2nd Edition). Redwood City, California: Addison-Wesley Publishing Co., 1991.

94. Xie, Lihua and Carlos E. de Souza. "H_∞ State Estimation for Linear Periodic Systems." *Proceedings of the 29th Conference on Decision and Control*. 3188–3193. December 1990.

95. Zarrop, Martin B. and Graham C. Goodwin. "Comments on 'Optimal Inputs for System Identification'," *IEEE Transactions on Automatic Control*, 299–300 (April 1975).

96. Zarrop, M.B. *Optimal Experiment Design for Dynamic System Identification*. Berlin: Springer-Verlag, 1979.

97. Zhang, Xue Jun. *Auxiliary Signal Design in Fault Detection and Diagnosis*, *134*. Lecture Notes in Control and Information Sciences. New York: Springer-Verlag, 1989. ISBN 0-387-51559-3.

Vita

Captain James Brown was born on 31 December 1958 in Louisiana, Missouri. He graduated from Louisiana High School in 1977 and enlisted in the United States Air Force in the Fall of that same year. After Air Force technical training at Chanute AFB, Illinois and Vandenberg AFB, California, Captain Brown served as a Titan II Missile Procedures Trainer Maintenance Specialist in the 390 Strategic Missile Wing at Davis-Monthan AFB, Arizona. After graduation from the University of Arizona under the Airman Education and Commissioning Program, he completed officer training and was commissioned in August of 1986. He went on to the Simulation and Modeling Directorate of Engineering, Munitions and Systems Division, Eglin AFB, Florida, where he supported the development of various weapon systems through simulation and systems analysis. Captain Brown's current assignment is with the Starfire Optical Range at Kirtland Air Force Base, New Mexico where he works as a controls engineer.

Permanent address: 9809 San Gabriel Rd NE
Albuquerque, NM 87111

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1995	Dissertation

4. TITLE AND SUBTITLE

Optimal Inputs for System Identification

5. FUNDING NUMBERS

6. AUTHOR(S)

James M. Brown II
Captain, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/DSG/95-S-04

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Peter K. Eide
WL/FIGS-2
Wright-Patterson AFB, OH 45433
(513) 255-8427

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The derivation of the power spectral density of the optimal input for system identification is addressed in this research. Optimality is defined in information theoretic terms, with entropy quantifying the parameter-information content of the input and output measurement sequences pertaining to a discrete-time plant. The maximization of entropy is performed in the context of three different scenarios. First, the case in which the average output power of the plant is constrained is considered. Second, input average power is constrained. Finally, the optimization is carried out unconstrained, but with penalties applied to both the input and output powers. Although the focus of this research is the enhancement of the parameter identification potential of general System Identification algorithms, a new and efficient System Identification algorithm that employs Iterated Weighted Least Squares is derived. Experimental evidence is presented which clearly illustrates the superiority of this algorithm. Furthermore, experiments are documented which corroborate and validate the maximum-entropy-based theory for optimal input design presented in this dissertation.

14. SUBJECT TERMS

system identification, parameter estimation

15. NUMBER OF PAGES

195

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL